LINEAR ALGEBRA

GABRIEL NAGY

Mathematics Department,
Michigan State University,
East Lansing, MI, 48824.

JULY 15, 2012

Abstract. These are the lecture notes for the course MTH 415, Applied Linear Algebra, a one semester class taught in 2009-2012. These notes present a basic introduction to linear algebra with emphasis on few applications. Chapter 1 introduces systems of linear equations, the Gauss-Jordan method to find solutions of these systems which transforms the augmented matrix associated with a linear system into reduced echelon form, where the solutions of the linear system are simple to obtain. We end the Chapter with two applications of linear systems: First, to find approximate solutions to differential equations using the method of finite differences; second, to solve linear systems using floating-point numbers, as happens in a computer. Chapter 2 reviews matrix algebra, that is, we introduce the linear combination of matrices, the multiplication of appropriate matrices, and the inverse of a square matrix. We end the Chapter with the LU-factorization of a matrix. Chapter 3 reviews the determinant of a square matrix, the relation between a non-zero determinant and the existence of the inverse matrix, a formula for the inverse matrix using the matrix of cofactors, and the Cramer rule for the formula of the solution of a linear system with an invertible matrix of coefficients. The advanced part of the course really starts in Chapter 4 with the definition of vector spaces, subspaces, the linear dependence or independence of a set of vectors, bases and dimensions of vector spaces. Both finite and infinite dimensional vector spaces are presented, however finite dimensional vector spaces are the main interest in this notes. Chapter 5 presents linear transformations between vector spaces, the components of a linear transformation in a basis, and the formulas for the change of basis for both vector components and transformation components. Chapter 6 introduces a new structure on a vector space, called an inner product. The definition of an inner product is based on the properties of the dot product in $\mathbb{R}^n$. We study the notion of orthogonal vectors, orthogonal projections, best approximations of a vector on a subspace, and the Gram-Schmidt orthonormalization procedure. The central application of these ideas is the method of least-squares to find approximate solutions to inconsistent linear systems. One application is to find the best polynomial fit to a curve on a plane. Chapter 8 introduces the notion of a normed space, which is a vector space with a norm function which does not necessarily comes from an inner product. We study the main properties of the p-norms on $\mathbb{R}^n$ or $\mathbb{C}^n$, which are useful norms in functional analysis. We briefly discuss induced operator norms. The last Section is an application of matrix norms. It discusses the condition number of a matrix and how to use this information to determine ill-conditioned linear systems. Finally, Chapter 9 introduces the notion of eigenvalue and eigenvector of a linear operator. We study diagonalizable operators, which are operators with diagonal matrix components in a basis of its eigenvectors. We also study functions of diagonalizable operators, with the exponential function as a main example. We also discuss how to apply these ideas to find solution of linear systems of ordinary differential equations.

Date: July 15, 2012, gnagy@math.msu.edu.

Table of Contents

Overview

Linear algebra is a collection of ideas involving algebraic systems of linear equations, vectors and vector spaces, and linear transformations between vector spaces.

Algebraic equations are called a system when there is more than one equation, and they are called linear when the unknown appears as a multiplicative factor with power zero or one. An example of a linear system of two equations in two unknowns is given in Eqs. (1.3)-(1.4) below. Systems of linear equations are the main subject of Chapter 1.

Examples of vectors are oriented segments on a line, plane, or space. An oriented segment is an ordered pair of points in these sets. Such ordered pair can be drawn as an arrow that starts on the first point and ends on the second point. Fix a preferred point in the line, plane or space, called the origin point, and then there exists a one-to-one correspondence between points in these sets and arrows that start at the origin point. The set of oriented segments with common origin in a line, plane, and space are called $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. A sketch of vectors in these sets can be seen in Fig. 1. Two operations are defined on oriented segments with common origin point: An oriented segment can be stretched or compressed; and two oriented segments with the same origin point can be added using the parallelogram law. An addition of several stretched or compressed vectors is called a linear combination. The set of all oriented segments with common origin point together with this operation of linear combination is the essential structure called vector space. The origin of the word "space" in the term "vector space" originates precisely in these examples, which were associated with the physical space.

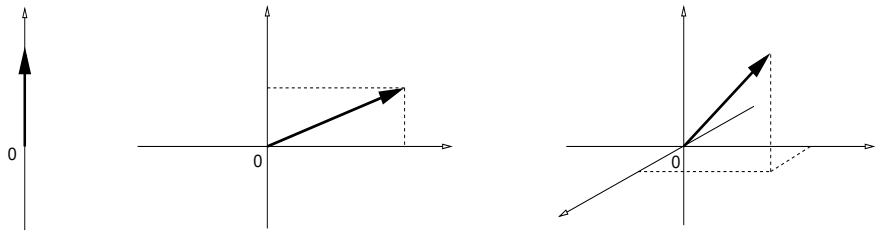

Figure 1. Example of vectors in the line, plane, and space, respectively.

Linear transformations are a particular type of functions between vector spaces that preserve the operation of linear combination. An example of a linear transformation is a 2×2 matrix $A = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ together with a matrix-vector product that specifies how this matrix transforms a vector on the plane into another vector on the plane. The result is thus a function $A : \mathbb{R}^2 \to \mathbb{R}^2$.

These notes try to be an elementary introduction to linear algebra with few applications.

Notation and conventions. We use the notation $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ to mean that $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Vectors will be denoted by boldface letters, like $\boldsymbol{u}$ and $\boldsymbol{v}$. The exception are a column vectors in $\mathbb{F}^n$ which are denoted in sanserif, like u and v. This notation permits to differentiate between a vector and its components on a basis. In a similar way, linear transformations between vector spaces are denoted by boldface capital letters, like $\boldsymbol{T}$ and $\boldsymbol{S}$. The exception are matrices in $\mathbb{F}^{m,n}$ which are denoted by capital sanserif letters like A and B. Again, this notation is useful to differentiate between a linear transformation and its

components on two bases. Below is a list of few mathematical symbols used in these notes:

$\mathbb{R}$	Set of real numbers,	$\mathbb{Q}$	Set of rational numbers,
$\mathbb{Z}$	Set of integer numbers,	$\mathbb{N}$	Set of positive integers,
$\{\boldsymbol{0}\}$	Zero set,	$\emptyset$	Empty set,
$\cup$	Union of sets,	$\cap$	Intersection of sets,
$:=$	Definition,	$\Rightarrow$	Implies,
$\forall$	For all,	$\exists$	There exists,
Proof	Beginning of a proof,	$\square$	End of a proof,
Example	Beginning of an example,	$\triangleleft$	End of an example.

Acknowledgments. I thanks all my students for pointing out several misprints and for helping make these notes more readable. I am specially grateful to Zhuo Wang and Wenning Feng.

CHAPTER 1. LINEAR SYSTEMS

1.1. ROW AND COLUMN PICTURES

1.1.1. **Row picture.** A central problem in linear algebra is to find solutions of a system of linear equations. A 2×2 *linear system* consists of two linear equations in two unknowns. More precisely, given the real numbers A_{11}, A_{12}, A_{21}, A_{22}, b_1, and b_2, find all numbers x and y solutions of both equations

$$A_{11}x + A_{12}y = b_1, \tag{1.1}$$

$$A_{21}x + A_{22}y = b_2. \tag{1.2}$$

These equations are called a system because there is more than one equation, and they are called linear because the unknowns, x and y, appear as multiplicative factors with power zero or one (for example, there is no term proportional to x^2 or to y^3). The *row picture* of a linear system is the method of finding solutions to this system as the intersection of all solutions to every single equation in the system. The individual equations are called *row equations*, or simply rows of the system.

EXAMPLE 1.1.1: Find all the numbers x and y solutions of the 2×2 linear system

$$2x - y = 0, \tag{1.3}$$

$$-x + 2y = 3. \tag{1.4}$$

SOLUTION: The solution to each row of the system above is found geometrically in Fig. 2.

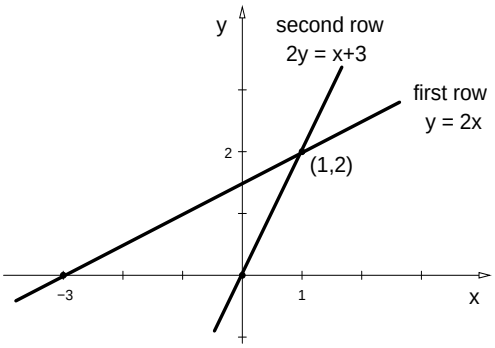

FIGURE 2. The solution of a 2×2 linear system in the row picture is the intersection of the two lines, which are the solutions of each row equation.

Analytically, the solution can be found by substitution:

$$2x - y = 0 \quad \Rightarrow \quad y = 2x \quad \Rightarrow \quad -x + 4x = 3 \quad \Rightarrow \quad \begin{cases} x = 1, \\ y = 2. \end{cases}$$

◁

An interesting property of the solutions to any 2×2 linear system is simple to prove using the row picture, and it is the following result.

Theorem 1.1.1. *Given any 2×2 linear system, only one of the following statements holds:*

(i) There exists a unique solution;
(ii) There exist infinitely many solutions;
(iii) There exists no solution.

It is interesting to remark what cannot happen, for example there is no 2×2 linear system having only two solutions. Unlike the quadratic equation $x^2 - 5x + 6 = 0$, which has two solutions given by $x = 2$ and $x = 3$, a 2×2 linear system has only one solution, or infinitely many solutions, or no solution at all. Examples of these three cases, respectively, are given in Fig. 3.

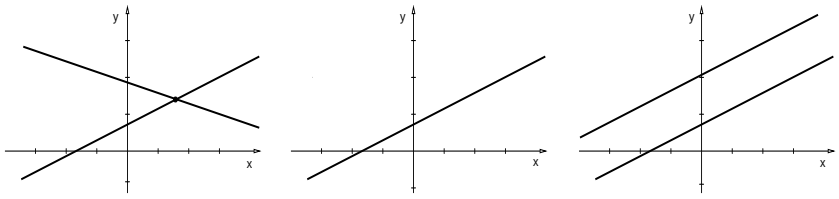

FIGURE 3. An example of the cases given in Theorem 1.1.1, cases *(i)-(iii)*.

Proof of Theorem 1.1.1: The solutions of each equation in a 2×2 linear system represents a line in $\mathbb{R}^2$. Two lines in $\mathbb{R}^2$ can intersect at a point, or can be coincident, or can be parallel but not coincident. These are the cases given in (i)-(iii). This establishes the Theorem. □

We now generalize the definition of a 2×2 linear system given in the Example 1.1.1 to m equations of n unknowns.

Definition 1.1.2. *An $m \times n$ **linear system** is a set of $m \geqslant 1$ linear equations in $n \geqslant 1$ unknowns is the following: Given the coefficients numbers A_{ij} and the source numbers b_i, with $i = 1, \cdots, m$ and $j = 1, \cdots n$, find the real numbers x_j solutions of*

$$A_{11}x_1 + \cdots + A_{1n}x_n = b_1$$
$$\vdots$$
$$A_{m1}x_1 + \cdots + A_{mn}x_n = b_m.$$

*Furthermore, an $m \times n$ linear system is called **consistent** iff it has a solution, and it is called **inconsistent** iff it has no solutions.*

EXAMPLE 1.1.2: Find all numbers x_1, x_2 and x_3 solutions of the 2×3 linear system

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 1 \\ -3x_1 - x_2 - 8x_3 &= 2 \end{aligned} \tag{1.5}$$

SOLUTION: Compute x_1 from the first equation, $x_1 = 1 - 2x_2 - x_3$, and substitute it in the second equation,

$$-3\,(1 - 2x_2 - x_3) - x_2 - 8x_3 = 2 \quad \Rightarrow \quad 5x_2 - 5x_3 = 5 \quad \Rightarrow \quad x_2 = 1 + x_3.$$

Substitute the expression for x_2 in the equation above for x_1, and we obtain

$$x_1 = 1 - 2\,(1 + x_3) - x_3 = 1 - 2 - 2x_3 - x_3 \quad \Rightarrow \quad x_1 = -1 - 3x_3.$$

Since there is no condition on x_3, the system above has infinitely many solutions parametrized by the number x_3. We conclude that $x_1 = -1 - 3x_3$, and $x_2 = 1 + x_3$, while x_3 is free. ◁

EXAMPLE 1.1.3: Find all numbers x_1, x_2 and x_3 solutions of the 3×3 linear system

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 2 \\ -x_1 + 2x_2 &= 1 \\ x_1 - x_2 + 2x_3 &= -2. \end{aligned} \tag{1.6}$$

Solution: While the row picture is appropriate to solve small systems of linear equations, it becomes difficult to carry out on 3×3 and bigger linear systems. The solution x_1, x_2, x_3 of the system above can be found as follows: Substitute the second equation into the first,

$$x_1 = -1 + 2x_2 \quad \Rightarrow \quad x_3 = 2 - 2x_1 - x_2 = 2 + 2 - 4x_2 - x_2 \quad \Rightarrow \quad x_3 = 4 - 5x_2;$$

then, substitute the second equation and $x_3 = 4 - 5x_2$ into the third equation,

$$(-1 + 2x_2) - x_2 + 2(4 - 5x_2) = -2 \quad \Rightarrow \quad x_2 = 1,$$

and then, substituting backwards, $x_1 = 1$ and $x_3 = -1$. We conclude that the solution is a single point in space given by $(1, 1, -1)$. ◁

The solution of each separate equation in the examples above represents a plane in $\mathbb{R}^3$. A solution to the whole system is a point that belongs to the three planes. In the 3×3 system in Example 1.1.3 above there is a unique solution, the point $(1, 1, -1)$, which means that the three planes intersect at a single point. In the general case, a 3×3 system can have a unique solution, infinitely many solutions or no solutions at all, depending on how the three planes in space intersect among them. The case with unique solution was represented in Fig. 4, while two possible situations corresponding to no solution are given in Fig. 5. Finally, two cases of 3×3 linear system having infinitely many solutions are pictured in Fig 6, where in the first case the solutions form a line, and in the second case the solutions form a plane because the three planes coincide.

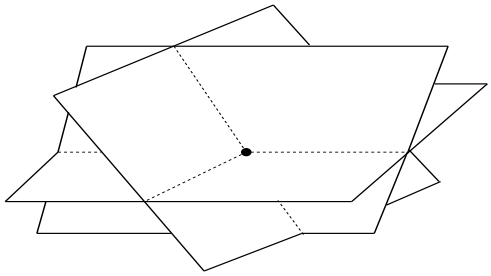

Figure 4. Planes representing the solutions of each row equation in a 3×3 linear system having a unique solution.

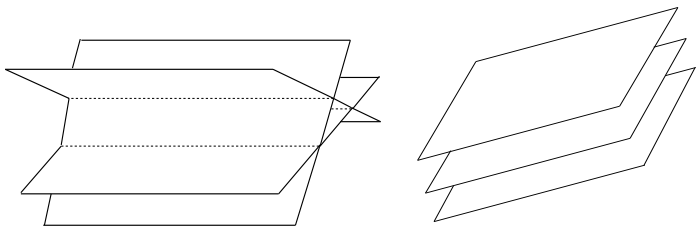

Figure 5. Two cases of planes representing the solutions of each row equation in 3×3 linear systems having no solutions.

Solutions of linear systems with more than three unknowns can not be represented in the three dimensional space. Besides, the substitution method becomes more involved to solve. As a consequence, alternative ideas are needed to solve such systems. We now discuss one

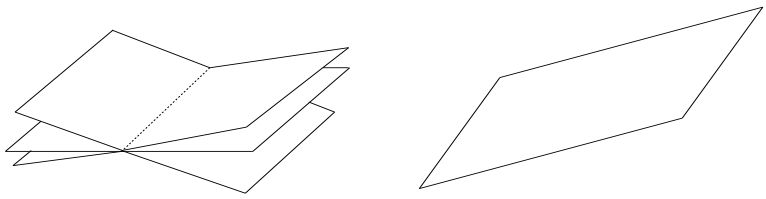

FIGURE 6. Two cases of planes representing the solutions of each row equation in 3×3 linear systems having infinitely many solutions.

of such ideas, the use of vectors to interpret and find solutions of linear systems. In the next Section we introduce another idea, the use of matrices and vectors to solve linear systems following the Gauss-Jordan method. This latter procedure is suitable to solve large systems of linear equations in an efficient way.

1.1.2. **Column picture.** Consider again the linear system in Eqs. (1.3)-(1.4) and introduce a change in the names of the unknowns, calling them x_1 and x_2 instead of x and y. The problem is to find the numbers x_1, and x_2 solutions of

$$2x_1 - x_2 = 0, \tag{1.7}$$

$$-x_1 + 2x_2 = 3. \tag{1.8}$$

We know that the answer is $x_1 = 1$, $x_2 = 2$. The row picture consisted in solving each row separately. The main idea in the column picture is to interpret the 2×2 linear system as an addition of new objects, column vectors, in the following way,

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} x_1 + \begin{bmatrix} -1 \\ 2 \end{bmatrix} x_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}. \tag{1.9}$$

The new objects are called column vectors and they are denoted as follows,

$$\mathsf{A}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}.$$

We can represent these vectors in the plane, as it is shown in Fig. 7.

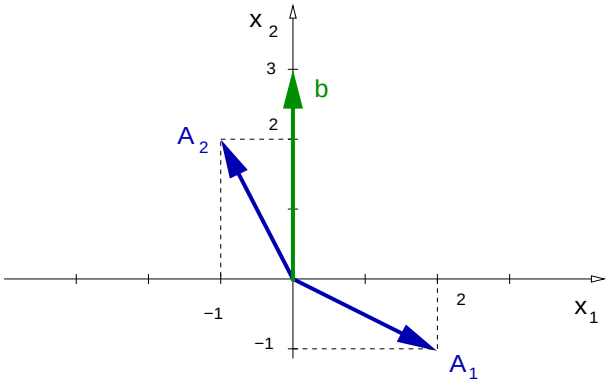

FIGURE 7. Graphical representation of column vectors in the plane.

The column vector interpretation of a 2×2 linear system determines an addition law of vectors and a multiplication law of a vector by a number. In the example above, we

know that the solution is given by $x_1 = 1$ and $x_2 = 2$, therefore in the column picture interpretation the following equation must hold

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} 2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}.$$

The study of this example suggests that the multiplication law of a vector by numbers and the addition law of two vectors can be defined by the following equations, respectively,

$$\begin{bmatrix} -1 \\ 2 \end{bmatrix} 2 = \begin{bmatrix} (-1)2 \\ (2)2 \end{bmatrix}, \qquad \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2-2 \\ -1+4 \end{bmatrix}.$$

The study of several examples of 2×2 linear systems in the column picture determines the following definition.

Definition 1.1.3. *The **linear combination** of the 2-vectors* $\mathsf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ *and* $\mathsf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$*, with the real numbers* a *and* b*, is defined as follows,*

$$a \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + b \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} au_1 + bv_1 \\ au_2 + bv_2 \end{bmatrix}.$$

A linear combination includes the particular cases of addition ($a = b = 1$), and multiplication of a vector by a number ($b = 0$), respectively given by,

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}, \qquad a \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} au_1 \\ au_2 \end{bmatrix}.$$

The addition law in terms of components is represented graphically by the parallelogram law, as it can be seen in Fig. 8. The multiplication of a vector by a number a affects the length and direction of the vector. The product $a\mathsf{u}$ stretches the vector u when $a > 1$ and it compresses u when $0 < a < 1$. If $a < 0$ then it reverses the direction of u and it stretches when $a < -1$ and compresses when $-1 < a < 0$. Fig. 8 represents some of these possibilities. Notice that the difference of two vectors is a particular case of the parallelogram law, as it can be seen in Fig. 9.

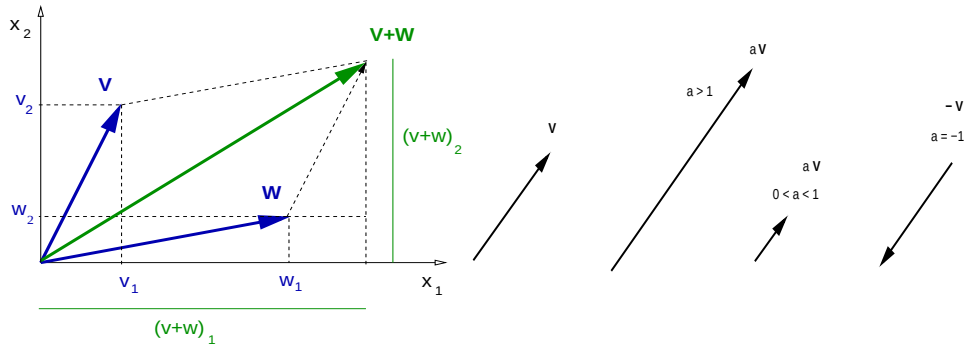

FIGURE 8. The addition of vectors can be computed with the parallelogram law. The multiplication of a vector by a number stretches or compresses the vector, and changes it direction in the case that the number is negative.

The *column picture* interpretation of a general 2×2 linear system given in Eqs. (1.1)-(1.2) is the following: Introduce the coefficient and source column vectors

$$\mathsf{A}_1 = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \tag{1.10}$$

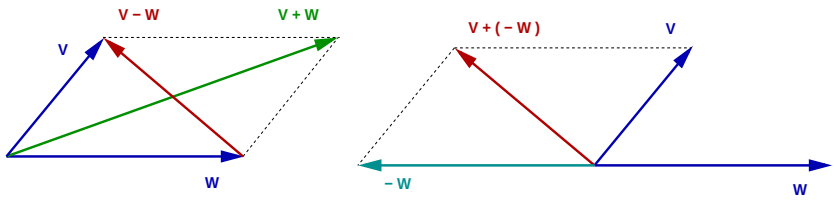

FIGURE 9. The difference of two vectors is a particular case of the parallelogram law of addition of vectors.

and then find the coefficients x_1 and x_2 that change the length of the coefficient column vectors A_1 and A_2 such that they add up to the source column vector b, that is,

$$\mathsf{A}_1\, x_1 + \mathsf{A}_2\, x_2 = \mathsf{b}.$$

For example, the column picture of the linear system in Eqs. (1.7)-(1.8) is given in Eq. (1.9). The solution of this system are the numbers $x_1 = 1$ and $x_2 = 2$, and this solution is represented in Fig. 10.

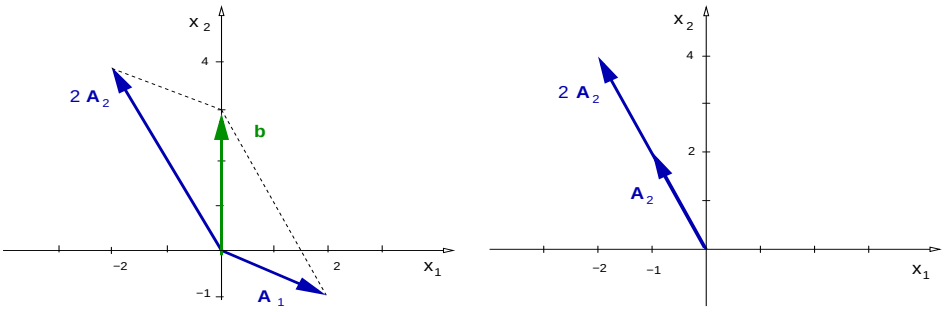

FIGURE 10. Representation of the solution of a 2×2 linear system in the column picture.

The existence and uniqueness of solutions in the case of 2×2 systems can be studied geometrically in the column picture as it was done in the row picture. In the latter case we have seen that all possible 2×2 systems fall into one of these three cases, unique solution, infinitely many solutions and no solutions at all. The proof was to study all possible ways two lines can intersect on the plane. The same existence and uniqueness statement can be proved in the column picture. In Fig. 11 we present these three cases in both row and column pictures. In the latter case the proof is to study all possible relative positions of the column vectors A_1, A_2, and b on the plane.

We see in Fig. 11 that the first case corresponds to a system with unique solution. There is only one linear combination of the coefficient vectors A_1 and A_2 which adds up to b. The reason is that the coefficient vectors are not proportional to each other. The second case corresponds to the case of infinitely many solutions. The coefficient vectors are proportional to each other and the source vector b is also proportional to them. So, there are infinitely many linear combinations of the coefficient vectors that add up to the source vector. The last case corresponds to the case of no solutions. While the coefficient vectors are proportional to each other, the source vector is not proportional to them. So, there is no linear combination of the coefficient vectors that add up to the source vector.

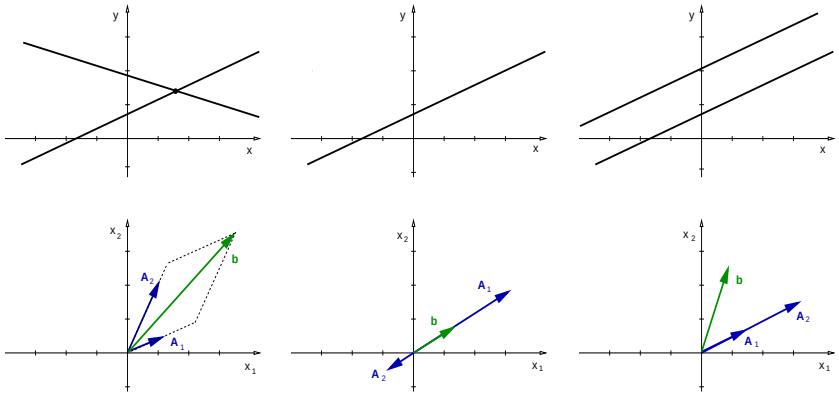

FIGURE 11. Examples of a solutions of general 2×2 linear systems having a unique, infinitely many, and no solution, represented in the row picture and in the column picture.

The ideas in the column picture can be generalized to $m \times n$ linear equations, which gives rise to the generalization to m-vectors of the definitions of linear combination presented above.

Definition 1.1.4. *The* ***linear combination*** *of the m-vectors* $\mathsf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$ *and* $\mathsf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$ *with the real numbers a, b is defined as follows*

$$a \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} + b \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} au_1 + bv_1 \\ \vdots \\ au_m + bv_m \end{bmatrix}.$$

This definition can be generalized to an arbitrary number of vectors. Column vectors provide a new way to denote an $m \times n$ system of linear equations.

Definition 1.1.5. *An $m \times n$* ***linear system*** *of $m \geqslant 1$ linear equations in $n \geqslant 1$ unknowns is the following: Given the coefficient m-vectors $\mathsf{A}_1, \cdots, \mathsf{A}_n$ and the source m-vector b, find the real numbers $x_1, \cdots, x_n$ solution of the linear combination*

$$\mathsf{A}_1 x_1 + \cdots + \mathsf{A}_n x_n = \mathsf{b}.$$

For example, recall the 3×3 system given as the second system in Eq. (1.6). This system in the column picture is the following: Find numbers x_1, x_2 and x_3 such that

$$\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} x_3 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}. \tag{1.11}$$

These are the main ideas in the column picture. We will see later that linear algebra emerges from the column picture. In the next Section we introduce the Gauss-Jordan method, which is a procedure to solve large systems of linear equations in an efficient way.

Further reading. For more details on the row picture see Section 1.1 in Lay's book [2]. There is a clear explanation of the column picture in Section 1.3 in Lay's book [2]. See also Section 1.2 in Strang's book [4] for a shorter summary of both the row and column pictures.

1.1.3. **Exercises.**

1.1.1.- Use the substitution method to find the solutions to the 2×2 linear system

$$2x - y = 1,$$
$$x + y = 5.$$

1.1.2.- Sketch the three lines solution of each row in the system

$$x + 2y = 2$$
$$x - y = 2$$
$$y = 1.$$

Is this linear system consistent?

1.1.3.- Sketch a graph representing the solutions of each row in the following **nonlinear** system, and decide whether it has solutions or not,

$$x^2 + y^2 = 4$$
$$x - y = 0.$$

1.1.4.- Graph on the plane the solution of each individual equation of the 3×2 linear system system

$$3x - y = 0,$$
$$x + 2y = 4,$$
$$-x + y = -2,$$

and determine whether the system is consistent or inconsistent.

1.1.5.- Show that the 3×3 linear system

$$x + y + z = 2,$$
$$x + 2y + 3z = 1,$$
$$y + 2z = 0,$$

is inconsistent, by finding a combination of the three equations that ads up to the equation $0 = 1$.

1.1.6.- Find all values of the constant k such that there exists infinitely many solutions to the 2×2 linear system

$$kx + 2y = 0,$$
$$x + \frac{k}{2}\, y = 0.$$

1.1.7.- Sketch a graph of the vectors

$$\mathsf{A}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Is the linear system $\mathsf{A}_1 x_1 + \mathsf{A}_2 x_2 = \mathsf{b}$ consistent? If the answer is "yes," find the solution.

1.1.8.- Consider the vectors

$$\mathsf{A}_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

(a) Graph the vectors A_1, A_2 and b on the plane.

(b) Is the linear system $\mathsf{A}_1 x_1 + \mathsf{A}_2 x_2 = \mathsf{b}$ consistent?

(c) Given the vector $\mathsf{c} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$, is the linear system $\mathsf{A}_1 x_1 + \mathsf{A}_2 x_2 = \mathsf{c}$ consistent? If the answer is "yes," is the solution unique?

1.1.9.- Consider the vectors

$$\mathsf{A}_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} -2 \\ -1 \end{bmatrix},$$

and given a real number h, set $\mathsf{c} = \begin{bmatrix} 2 \\ h \end{bmatrix}$. Find all values of h such that the system $\mathsf{A}_1 x_1 + \mathsf{A}_2 x_2 = \mathsf{c}$ is consistent.

1.1.10.- Show that the three vectors below lie on the same plane, by expressing the third vector as a linear combination of the first two, where

$$\mathsf{A}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \mathsf{A}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \ \mathsf{A}_3 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}.$$

Is the linear system

$$\mathsf{A}_1 x_1 + \mathsf{A}_2 x_2 + \mathsf{A}_3 x_3 = \mathsf{0}$$

consistent? If the answer is "yes," is the solution unique?

1.2. Gauss-Jordan method

1.2.1. **The augmented matrix.** Solutions to $m \times n$ linear systems can be obtained in an efficient way using the Gauss-Jordan method. Efficient here means to perform as few algebraic operations as possible either to find the solution or to show that the solution does not exist. Before introducing this method, we need several definitions.

Definition 1.2.1. *The* ***coefficients matrix****, the* ***source vector****, and the* ***augmented matrix*** *of the $m \times n$ linear system*

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= b_1 \\ &\vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= b_m, \end{aligned}$$

are given by, respectively,

$$\mathsf{A} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad [\mathsf{A}|\mathsf{b}] = \left[\begin{array}{ccc|c} A_{11} & \cdots & A_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ A_{m1} & \cdots & A_{mn} & b_m \end{array}\right].$$

We call A an $m \times n$ matrix, with m the number of rows and n the number of columns. The source vector b is a particular case of an $m \times 1$ matrix. The augmented matrix of an $m \times n$ linear system is given by the coefficient matrix and the source vector together, hence it is an $m \times (n+1)$ matrix.

Example 1.2.1: Find the coefficient matrix, the source vector and the augmented matrix of the 2×2 linear system

$$2x_1 - x_2 = 0, \tag{1.12}$$
$$-x_1 + 2x_2 = 3. \tag{1.13}$$

Solution: The coefficient matrix is 2×2, the source vector is 2×1, and the augmented matrix is 2×3, given respectively by

$$\mathsf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad [\mathsf{A}|\mathsf{b}] = \left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & 3 \end{array}\right]. \tag{1.14}$$

◁

Example 1.2.2: Find the coefficient matrix, the source vector and the augmented matrix of the 2×3 linear system

$$\begin{aligned} 2x_1 - x_2 &= 0, \\ -x_1 + 2x_2 + 3x_3 &= 0. \end{aligned}$$

Solution: The coefficient matrix is 2×3, the source vector is 2×1, and the augmented matrix is 2×4, given respectively by

$$\mathsf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 3 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad [\mathsf{A}|\mathsf{b}] = \left[\begin{array}{ccc|c} 2 & -1 & 0 & 0 \\ -1 & 2 & 3 & 0 \end{array}\right].$$

Notice that the coefficent matrix in this example is equal to the augmented matrix in the previous Example. This means that the vertical separator in the definition of the augmented

matrix is important. If one does not write down the vertical separator in Example 1.2.1, then one is actually working on the system in Example 1.2.2. ◁

We also use the alternative notation $\mathsf{A} = [A_{ij}]$ to denote a matrix with components A_{ij}, and $\mathsf{b} = [b_i]$ to denote a vector with components b_i, where $i = 1, \cdots, m$ and $j = 1, \cdots, n$.

Definition 1.2.2. *The **diagonal** of an $m \times n$ matrix $\mathsf{A} = [A_{ij}]$ is the set of all coefficients A_{ii} for $i = 1, \cdots, m$. A matrix is **upper triangular** iff every coefficient below the diagonal vanishes, and **lower triangular** iff every coefficient above the diagonal vanishes.*

EXAMPLE 1.2.3: Highlight the diagonal coefficients in 3×3, 2×3 and 3×2 matrices.

SOLUTION: The diagonal coefficients in the 3×3, 2×3 and 3×2 matrices below are highlighted in green, where the coefficients with $*$ denote non-diagonal coefficients:

$$\begin{bmatrix} A_{11} & * & * \\ * & A_{22} & * \\ * & * & A_{33} \end{bmatrix}, \quad \begin{bmatrix} A_{11} & * & * \\ * & A_{22} & * \end{bmatrix}, \quad \begin{bmatrix} A_{11} & * \\ * & A_{22} \\ * & * \end{bmatrix}.$$

◁

EXAMPLE 1.2.4: Write down the most general 3×3, 2×3 and 3×2 upper triangular matrices, highlighting the diagonal elements.

SOLUTION: The following matrices are upper triangular:

$$\begin{bmatrix} A_{11} & * & * \\ 0 & A_{22} & * \\ 0 & 0 & A_{33} \end{bmatrix}, \quad \begin{bmatrix} A_{11} & * & * \\ 0 & A_{22} & * \end{bmatrix}, \quad \begin{bmatrix} A_{11} & * \\ 0 & A_{22} \\ 0 & 0 \end{bmatrix}.$$

◁

1.2.2. **Gauss elimination operations.** The Gauss-Jordan Method is a procedure performed on the augmented matrix of a linear system. It consists on a sequence of operations, called Gauss elimination operations, that change the augmented matrix of the system but they do not change the solutions of the system. The Gauss elimination operations were already known in China around 200 BC. We call them after Carl Friedrich Gauss, since he made them very popular around 1810, when he used them to study the orbit of the asteroid Pallas, giving a systematic method to solve a 6×6 algebraic linear system.

Definition 1.2.3. *The **Gauss elimination operations** are three operations on a matrix:*

(i) To add a multiple of one row to another row;
(ii) To interchange two rows;
(iii) To multiply a row by a non-zero number.

These operations are respectively represented by the symbols given in Fig. 12.

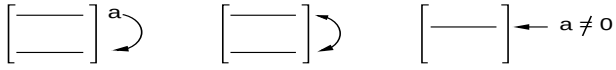

FIGURE 12. A representation of the Gauss elimination operations (i), (ii) and (iii), respectively.

REMARK: If the factor in operation (iii) is allowed to be zero, then multiplying an equation by zero modifies the solutions of the original system, since such operation on an equation is equivalent to erase that equation from the system. This explains why the factor in operation (iii) is constrained to be non-zero.

The Gauss elimination operations change the coefficients of the augmented matrix of a system but do not change its solution. Two systems of linear equations having the same solutions are called *equivalent*. The Gauss-Jordan Method is an algorithm using these operations that transforms any linear system into an equivalent system where the solutions are given explicitly. We describe the Gauss-Jordan Method only through examples.

Example 1.2.5: Find the solution of the 2×2 linear system with augmented matrix given in Eq. (1.14) using Gauss operations.

Solution:

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & 3 \end{array}\right] \to \left[\begin{array}{cc|c} 2 & -1 & 0 \\ -2 & 4 & 6 \end{array}\right] \to \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 3 & 6 \end{array}\right] \to$$

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 1 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 2 & 0 & 2 \\ 0 & 1 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

The Gauss operations have changed the augmented matrix of the original system as follows:

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & 3 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

Since the Gauss operations do not change the solution of the associated linear systems, the augmented matrices above imply that the following two linear systems have the same solutions,

$$\left.\begin{aligned} 2x_1 - x_2 &= 0, \\ -x_1 + 2x_2 &= 3. \end{aligned}\right\} \quad \Leftrightarrow \quad \begin{cases} x_1 = 1, \\ x_2 = 2. \end{cases}$$

On the last system the solution is given explicitly, $x_1 = 1$, $x_2 = 2$, since no additional algebraic operations are needed to find the solution. ◁

A precise way to define the notion of an augmented matrix corresponding to a linear system with solutions "easy to read" is captured in the notions of echelon form of a matrix, and reduced echelon form of a matrix. Next Section is dedicated to present these notions in some detail.

1.2.3. **Square systems.** In the rest of this Section we study a particular type of linear systems having the same number of equations and unknowns.

Definition 1.2.4. *An $m \times n$ linear system is called a* ***square system*** *iff holds that $m = n$.*

An example of a square system is the 2×2 linear system in Eqs. (1.12)-(1.13). In the rest of this Section we introduce the Gauss operations and back substitution method to find solutions to square linear systems. We later on compare this method with the Gauss-Jordan Method restricted to square linear systems.

We start using *Gauss operations and back substitution* to find solutions to square $n \times n$ linear systems. This method has two main parts: First, use Gauss operations to transform the augmented matrix of the system into an upper triangular form. Second, use back substitution to compute the solution to the system.

Example 1.2.6: Use the Gauss operations and back substitution to solve the 3×3 system

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 2 \\ -x_1 + 2x_2 &= 1 \\ x_1 - x_2 + 2x_3 &= -2. \end{aligned}$$

Solution: We have already seen in Example 1.1.3 that the solution of this system is given by $x_1 = x_2 = 1$, $x_3 = -1$. Let us find that solution using Gauss operations with back

substitution. First transform the augmented matrix of the system into upper triangular form:

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 2 \\ -1 & 2 & 0 & 1 \\ 1 & -1 & 2 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & 2 & -2 \\ -1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & 2 & -2 \\ 0 & 1 & 2 & -1 \\ 0 & 3 & -3 & 6 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & -1 & 2 & -2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & -9 & 9 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & 2 & -2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right].$$

We now write the linear system corresponding to the last augmented matrix,

$$\begin{aligned} 2x_1 - x_2 + 2x_3 &= -2 \\ x_2 + 2x_3 &= -1 \\ x_3 &= -1. \end{aligned}$$

We now use the back substitution to obtain the solution, that is, introduce $x_3 = -1$ into the second equation, which gives us $x_2 = 1$. Then, substitute $x_3 = -1$ and $x_2 = 1$ into the first equation, which gives us $x_1 = 1$. ◁

We now use the *Gauss-Jordan Method* to find solutions to square $n \times n$ linear systems. This is a minor modification of the Gauss operations and back substitution method. The difference is that now we do not stop doing Gauss operations when the augmented matrix becomes upper triangular. We keep doing Gauss operations in order to make zeros above the diagonal. Then, back substitution will no be needed to find the solution of the linear system. The solution will be given explicitly at the end of the procedure.

EXAMPLE 1.2.7: Use the Gauss-Jordan method to solve the same 3×3 linear system as in Example 1.2.6, that is,

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 2 \\ -x_1 + 2x_2 &= 1 \\ x_1 - x_2 + 2x_3 &= -2. \end{aligned}$$

SOLUTION: In Example 1.2.6 we performed Gauss operations on the augmented matrix of the system above until we obtained an upper triangular matrix, that is,

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 2 \\ -1 & 2 & 0 & 1 \\ 1 & -1 & 2 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & 2 & -2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right].$$

The idea now is to continue with Gauss operations, as follows:

$$\left[\begin{array}{ccc|c} 1 & -1 & 2 & -2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 4 & -3 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right] \quad \Rightarrow \quad \begin{cases} x_1 = 1, \\ x_2 = 1, \\ x_3 = -1. \end{cases}$$

In the last step we do not need to do back substitution to compute the solution. It is obtained without doing any further algebraic operation. ◁

Further reading. Almost every linear algebra book describes the Gauss-Jordan method. See Section 1.2 in Lay's book [2] for a summary of echelon forms and the Gauss-Jordan method. See Sections 1.2 and 1.3 in Meyer's book [3] for more details on Gauss elimination operations and back substitution. Also see Section 1.3 in Strang's book [4].

1.2.4. **Exercises.**

1.2.1.- Use Gauss operations and back substitution to find the solution of the 3×3 linear system

$$\begin{aligned} x_1 + x_2 + x_3 &= 1, \\ x_1 + 2x_2 + 2x_3 &= 1, \\ x_1 + 2x_2 + 3x_3 &= 1. \end{aligned}$$

1.2.2.- Use Gauss operations and back substitution to find the solution of the 3×3 linear system

$$\begin{aligned} 2x_1 - 3x_2 &= 3, \\ 4x_1 - 5x_2 + x_3 &= 7, \\ 2x_1 - x_2 - 3x_3 &= 5. \end{aligned}$$

1.2.3.- Find the solution of the following linear system with Gauss-Jordan's method,

$$\begin{aligned} 4x_2 - 3x_3 &= 3, \\ -x_1 + 7x_2 - 5x_3 &= 4, \\ -x_1 + 8x_2 - 6x_3 &= 5. \end{aligned}$$

1.2.4.- Find the solutions to the following two linear systems, which have the same matrix of coefficient A but different source vectors b_1 and b_2, given respectively by,

$$\begin{aligned} 4x_1 - 8x_2 + 5x_3 &= 1, & 4x_1 - 8x_2 + 5x_3 &= 0, \\ 4x_1 - 7x_2 + 4x_3 &= 0, & 4x_1 - 7x_2 + 4x_3 &= 1, \\ 3x_1 - 4x_2 + 2x_3 &= 0, & 3x_1 - 4x_2 + 2x_3 &= 0. \end{aligned}$$

Solve these two systems at one time using the Gauss-Jordan method on an augmented matrix of the form $[\mathsf{A}|\mathsf{b}_1|\mathsf{b}_2]$.

1.2.5.- Use the Gauss-Jordan method to solve the following three linear systems at the same time,

$$\begin{aligned} 2x_1 - x_2 &= 1, & &= 0, & &= 0, \\ -x_1 + 2x_2 - x_3 &= 0, & &= 1, & &= 0, \\ -x_2 + x_3 &= 0, & &= 0, & &= 1. \end{aligned}$$

1.3. Echelon forms

1.3.1. **Echelon and reduced echelon forms.** The Gauss-Jordan method is a procedure that uses Gauss operations to transform the augmented matrix of an $m \times n$ linear system into the augmented matrix of an equivalent system whose solutions can be found without performing further algebraic operations. A precise way to define the notion of a linear system with solutions that can be found without doing further algebraic operations is captured in the notions of echelon form and reduced echelon form of its augmented matrix.

Definition 1.3.1. *An $m \times n$ matrix is in **echelon form** iff the following conditions hold:*

(i) The zero rows are located at the bottom rows of the matrix;
(ii) The first non-zero coefficient on a row is always to the right of the first non-zero coefficient of the row above it.

*The **pivot** coefficient is the first non-zero coefficient on every non-zero row in a matrix in echelon form.*

Example 1.3.1: The 6×8, 3×5 and 3×3 matrices given below are in echelon form, where pivots are highlighted and the $*$ means any non-zero number.

$$\begin{bmatrix} \mathbf{*} & * & * & * & * & * & * & * \\ 0 & 0 & \mathbf{*} & * & * & * & * & * \\ 0 & 0 & 0 & \mathbf{*} & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{*} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{*} & * & * & * & * \\ 0 & 0 & \mathbf{*} & * & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{*} & * & * \\ 0 & \mathbf{*} & * \\ 0 & 0 & \mathbf{*} \end{bmatrix}.$$

◁

Example 1.3.2: The following matrices are in echelon form, with pivots highlighted:

$$\begin{bmatrix} \mathbf{1} & 3 \\ 0 & \mathbf{1} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{2} & 3 & 2 \\ 0 & \mathbf{4} & -2 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{2} & 1 & 1 \\ 0 & \mathbf{3} & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

◁

Definition 1.3.2. *An $m \times n$ matrix is in **reduced echelon form** iff the matrix is in echelon form and the following two conditions hold:*

(i) The pivot coefficient is equal to 1;
(ii) The pivot coefficient is the only non-zero coefficient in that column.

We denote by E_A a reduced echelon form of a matrix A.

Example 1.3.3: The 6×8, 3×5 and 3×3 matrices given below are in echelon form, where pivots are highlighted and the $*$ means any non-zero number.

$$\begin{bmatrix} \mathbf{1} & * & 0 & 0 & * & * & 0 & * \\ 0 & 0 & \mathbf{1} & 0 & * & * & 0 & * \\ 0 & 0 & 0 & \mathbf{1} & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{1} & * & 0 & * & * \\ 0 & 0 & \mathbf{1} & * & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \end{bmatrix}.$$

◁

Example 1.3.4: And the following matrices are not only in echelon form but also in reduced echelon form; again, pivot coefficients are highlighted:

$$\begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{1} & 0 & 4 \\ 0 & \mathbf{1} & 5 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

◁

Summarizing, the Gauss-Jordan method uses Gauss operations to transform the augmented matrix of a linear system into reduced echelon form. Then, the solutions of the linear system can be obtained without doing further algebraic operations.

Example 1.3.5: Consider a 3×3 linear system with augmented matrix having a reduced echelon form given below. Then, the solution of this linear system is simple to obtain:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad \Rightarrow \quad \begin{cases} x_1 = -1 \\ x_2 = 3 \\ x_3 = 2. \end{cases}$$

◁

We use the notation $\mathsf{E_A}$ for a reduced echelon form of an $m \times n$ matrix A, since a reduced echelon form of a matrix has an important property: It is unique. We state this property in Proposition 1.3.3 below. Since the proof is somehow involved, after the proof we show a different proof for the particular case of 2×2 matrices.

Proposition 1.3.3. *The reduced echelon $\mathsf{E_A}$ form of an $m \times n$ matrix A is unique.*

Proof of Proposition 1.3.3: We assume that there exists an $m \times n$ matrix A having two reduced echelon forms $\mathsf{E_A}$ and $\hat{\mathsf{E}}_\mathsf{A}$, and we will show that $\mathsf{E_A} = \hat{\mathsf{E}}_\mathsf{A}$. Introduce the column vector notation for matrices (see Sect. 1.4) and for vectors (see Sect. 1.4), respectively,

$$\mathsf{A} = \big[\mathsf{A}_1, \cdots, \mathsf{A}_n\big], \qquad \mathsf{E_A} = \big[\mathsf{E}_1, \cdots, \mathsf{E}_n\big], \qquad \hat{\mathsf{E}}_\mathsf{A} = \big[\hat{\mathsf{E}}_1, \cdots, \hat{\mathsf{E}}_n\big], \qquad \mathsf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

If the reduced echelon forms $\mathsf{E_A}$ and $\hat{\mathsf{E}}_\mathsf{A}$ were different, then they should start differing at a particular column, say column i, with $1 \leqslant i \leqslant n$. That is, there would be vectors E_i and $\hat{\mathsf{E}}_i$ satisfying $\mathsf{E}_i \neq \hat{\mathsf{E}}_i$ and $\mathsf{E}_j = \hat{\mathsf{E}}_j$ for $j = 1, \cdots, (i-1)$. The proof of Proposition 1.3.3 reduces to show that this is not the case, by studying the following two cases:

(a) The vector of one of the reduced echelon form matrices, say the column vector E_i of $\mathsf{E_A}$, does not contain a pivot coefficient.

(b) Both vectors E_i and $\hat{\mathsf{E}}_i$ do contain a pivot coefficient.

Consider the case in (a). The definition of reduced echelon form implies that the vector $\mathsf{E}_i = [e_j]$ has non-zero components among the first k components, with $0 \leqslant k < i$, hence, $e_j = 0$ for $j = (k+1), \cdots, n$. Furthermore, the definition of reduced echelon form also implies that there are k columns to the left of vector E_i with pivot coefficients. Let us denote these columns by E_{p_j}, for $j = 1, \cdots, k$. We use a new subindex p_j, since the k columns need not be the first k columns. This information about the reduced echelon form matrix $\mathsf{E_A}$ can be translated into a vector x solution of the linear system with augmented matrix $\big[\mathsf{E_A}\big|\mathsf{0}\big]$. This solution is the vector $\mathsf{x} = [x_j]$ with non-zero components $x_{p_j} = e_j$ for $j = 1, \cdots, k$, and $x_i = -1$, while the rest of the components vanish, since the following equations hold,

$$e_1\mathsf{E}_{p_1} + \cdots + e_k\mathsf{E}_{p_k} + (-1)\mathsf{E}_i = \mathsf{0} \quad \Leftrightarrow \quad x_{p_1}\mathsf{E}_{p_1} + \cdots + x_{p_k}\mathsf{E}_{p_k} + x_i\mathsf{E}_i = \mathsf{0}.$$

Recall now that the solution of a linear system is not changed when Gauss operations are performed on the coefficient matrix. Since Gauss operations change $\mathsf{E_A} \to \mathsf{A} \to \hat{\mathsf{E}}_\mathsf{A}$, the same vector x above is solution of the linear system with augmented matrix $\left[\hat{\mathsf{E}}_\mathsf{A}|\mathsf{0}\right]$, that is,

$$x_{p_1}\hat{\mathsf{E}}_{p_1} + \cdots + x_{p_k}\hat{\mathsf{E}}_{p_k} + x_i\hat{\mathsf{E}}_i = \mathsf{0} \quad \Leftrightarrow \quad e_1\hat{\mathsf{E}}_{p_1} + \cdots + e_k\hat{\mathsf{E}}_{p_k} + (-1)\hat{\mathsf{E}}_i = \mathsf{0}.$$

Since $\mathsf{E}_j = \hat{\mathsf{E}}_j$ for $j = 1, \cdots, k$, the second equation above implies that $\mathsf{E}_i = \hat{\mathsf{E}}_i$ also holds. We conclude that the first vector in $\mathsf{E_A}$ that differs from a vector in $\hat{\mathsf{E}}_\mathsf{A}$ cannot be a vector without pivot coefficients.

Consider the case in (b), that is, if E_i and $\hat{\mathsf{E}}_i$ are the first vectors from $\mathsf{E_A}$ and $\hat{\mathsf{E}}_\mathsf{A}$, respectively, which are different, then both vectors have pivot coefficients. (If only one of these vectors had a pivot coefficient, then the argument in case (a) on the vector without pivot coefficient would imply that the first vector could not have a pivot coefficient.) Suppose that the vector $\mathsf{E}_i = [e_j]$ has the pivot coefficient at the position k_0, that is, $e_j = 0$ for $j \neq k_)$ and $e_{k_0} = 1$. Similarly, suppose that the vector $\hat{\mathsf{E}}_i = [\hat{e}_j]$ has the pivot coefficient at the position k_1, that is, $\hat{e}_j = 0$ for $j \neq k_)$ and $\hat{e}_{k_1} = 1$. By definition of reduced echelon form all rows in $\mathsf{E_A}$ above the k_0 row have pivots columns to the left of E_i. This statement also applies to matrix $\hat{\mathsf{E}}_\mathsf{A}$, since $\hat{\mathsf{E}}_i$ is the first column that differs from $\mathsf{E_A}$. Therefore $k_0 \leqslant k_1$. The exactly same argument also applies to matrix $\hat{\mathsf{E}}_\mathsf{A}$ concluding that $k_1 \leqslant k_0$. Therefore, $k_0 = k_1$, and so, $\mathsf{E}_i = \hat{\mathsf{E}}_i$. We conclude that the first vector in $\mathsf{E_A}$ that differs from a vector in $\hat{\mathsf{E}}_\mathsf{A}$ cannot be a vector with pivot a coefficient. Therefore, parts (a) and (b) establish the Proposition. □

The proof above is not straightforward to understand, so it might be a good idea to present a simpler proof for the particular case of 2×2 matrices.

Alternative proof of Proposition 1.3.3 in the case of 2×2 matrices: We start recalling once again the following observation that holds for any $m \times n$ matrix A: Since Gauss operations do not change the solutions of the homogeneous system $[\mathsf{A}|\mathsf{0}]$, if a matrix A has two different reduced echelon forms, $\mathsf{E_A}$ and $\tilde{\mathsf{E}}_\mathsf{A}$, then the set of solutions of the systems $[\mathsf{E_A}|\mathsf{0}]$ and $[\tilde{\mathsf{E}}_\mathsf{A}|\mathsf{0}]$ must coincide with the solutions of the system $[\mathsf{A}|\mathsf{0}]$. What we are going to show now is the following: All possible 2×2 reduced echelon form matrices E have different solutions to the homogeneous equation $[\mathsf{E}|\mathsf{0}]$. This property then establishes that every 2×2 matrix A has a unique reduced echelon form.

Given any 2×2 matrix A, all possible reduced echelon forms are the following:

$$\mathsf{E_A} = \begin{bmatrix} 1 & c \\ 0 & 0 \end{bmatrix}, \quad c \in \mathbb{R}, \quad \mathsf{E_A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathsf{E_A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We claim that all these matrices determine different sets of solutions to the equation $[\mathsf{E_A}|\boldsymbol{0}]$. In the first case, the set of solutions are lines given by

$$x_1 = -c\,x_2, \quad c \in \mathbb{R}. \tag{1.15}$$

In the second case, the set of solutions is a single line given by

$$x_2 = 0, \quad x_1 \in \mathbb{R}.$$

Notice that this line does not belong to the set given in Eq. (1.15). In the third case, the solution is a single point

$$x_1 = 0, \quad x_2 = 0.$$

Since all these sets of solutions are different, and only one corresponds to the equation $[\mathsf{A}|\mathsf{0}]$, then there is only one reduced echelon form $\mathsf{E_A}$ for matrix A. This establishes the Proposition for 2×2 matrices. □

While the reduced echelon form of a matrix is unique, a matrix may have many different echelon forms. Also notice that given a matrix A, there are many different sequences of Gauss operations that produce the reduced echelon form E_A, as it is sketched in Fig. 13.

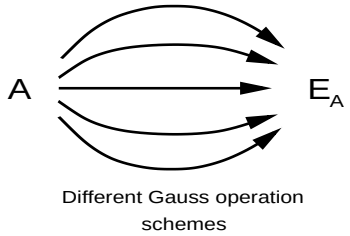

FIGURE 13. The reduced echelon form of a matrix can be obtained in many different ways.

EXAMPLE 1.3.6: Use two different sequences of Gauss operations to find the reduced echelon form of matrix

$$\mathsf{A} = \begin{bmatrix} 2 & 4 & 10 \\ 1 & 3 & 7 \end{bmatrix}$$

SOLUTION: Here are two different sequences of Gauss operations to find E_A:

$$\mathsf{A} = \begin{bmatrix} 2 & 4 & 10 \\ 1 & 3 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 5 \\ 1 & 3 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \mathsf{E}_\mathsf{A},$$

$$\mathsf{A} = \begin{bmatrix} 2 & 4 & 10 \\ 1 & 3 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 7 \\ 2 & 4 & 10 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 5 \\ 0 & -2 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \mathsf{E}_\mathsf{A}.$$

The matrix E_A is the same using the two Gauss operation sequences above. ◁

1.3.2. **The rank of a matrix.** Since the reduced echelon form E_A of an $m \times n$ matrix A is unique, any property of matrix E_A is indeed a property of matrix A. In particular, the uniqueness of E_A implies that the number of pivots of E_A is also unique. This number is a property of matrix A, and we give that number a name, since it will be important later on.

Definition 1.3.4. *The **rank** of an $m \times n$ matrix A, denoted as* rank(A)*, is the number of pivots in its reduced echelon form E_A.*

EXAMPLE 1.3.7: Find the rank of the coefficient matrix and all solutions of the linear system

$$\begin{aligned} x_1 + 3x_3 &= -1 \\ x_2 - 2x_3 &= 3 \\ 2x_1 + x_2 + 4x_3 &= 1. \end{aligned}$$

SOLUTION: The 3×3 system has an augmented matrix with reduced echelon form

$$\left[\begin{array}{ccc|c} 1 & 0 & 3 & -1 \\ 0 & 1 & -2 & 3 \\ 2 & 1 & 4 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 3 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Since the reduced echelon form of the coefficient matrix has two pivots, we obtain that the coefficient matrix has rank two. The solutions of the linear system can be obtained without

any further calculations, since the coefficient matrix is already in reduced echelon form. These solutions are the following,

$$\left[\begin{array}{ccc|c} 1 & 0 & 3 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \Rightarrow \quad \begin{cases} x_1 = -1 - 3x_3, \\ x_2 = 3 + 2x_3, \\ x_3 : \text{ free variable.} \end{cases}$$

We call x_3 a free variable, since there is a solution for every value of this variable. ◁

Definition 1.3.5. *A variable in a linear system is called a* ***free variable*** *iff there is a solution of the linear system for every value of that variable.*

In Example 1.3.7 the free variable is x_3, since for every value of x_3 the other two variables x_1 and x_2 are fixed by the system. Notice that *only the number of free variables is relevant*, but not which particular variable is the free one. In the following example we express the solutions in Example 1.3.7 in terms of x_1 as a free variable.

EXAMPLE 1.3.8: Express the solutions in Example 1.3.7 in terms of x_1 as a free variable.

SOLUTION: The solutions given in Example 1.3.7 can also be expressed as

$$\begin{aligned} x_1 &: \text{ free variable}, \\ x_2 &= 3 + \frac{2}{3}(-1 - x_1), \\ x_3 &= \frac{1}{3}(-1 - x_1). \end{aligned}$$

◁

As the reader may have noticed there is a relation between the rank of the coefficient matrix and the number of free variables in the solution of the linear system.

Theorem 1.3.6. *The number of free variables in the solutions of an $m \times n$ consistent linear system with augmented matrix $[\mathsf{A}|\mathsf{b}]$ is given by $\big(n - \text{rank}(\mathsf{A})\big)$.*

Proof of Theorem 1.3.6: The number of non-pivots columns in $\mathsf{E_A}$ is actually the number of variables not fixed by the linear system with augmented matrix $[\mathsf{A}|\mathsf{b}]$. The number of non-pivot columns is the total number of columns minus the pivot columns, that is, $\big(n - \text{rank}(A)\big)$. This establishes the Theorem. □

We saw in Example 1.3.7 that the 3×3 coefficient matrix has $\text{rank}(\mathsf{A}) = 2$. Since the system has $n = 3$ variables, we conclude, without actually computing the solution, that the solution has $n - \text{rank}(\mathsf{A}) = 3 = 2 = 1$ free variable. The free variable can be any variable in the system. The relevant information is that there is one free variable.

EXAMPLE 1.3.9: We now present three examples of consistent linear systems.

(i) Show that the 2×2 linear system below is consistent with coefficient matrix having rank one and solutions having one free variable,

$$\begin{aligned} 2x_1 - x_2 &= 1 \\ -\frac{1}{2}x_1 + \frac{1}{4}x_1 &= -\frac{1}{4}. \end{aligned}$$

SOLUTION: Gauss operations transform the augmented matrix of the system as follows,

$$\left[\begin{array}{cc|c} 2 & -1 & 1 \\ -1/2 & 1/4 & -1/4 \end{array}\right] \to \left[\begin{array}{cc|c} 2 & -1 & 1 \\ 0 & 0 & 0 \end{array}\right].$$

The system is consistent, has rank one, and has a free variable, hence the system has infinitely many solutions.

(ii) Show that the 2×3 linear system below is consistent with coefficient matrix having rank two and solution having one free variable,

$$x_1 + 2x_2 + 3x_3 = 1,$$
$$3x_1 + 5x_2 + 2x_3 = 2.$$

Solution: Perform the following Gauss operations on the augmented matrix,

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 5 & 2 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -1 & -7 & -1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -11 & -1 \\ 0 & 1 & 7 & 1 \end{array}\right].$$

We see that the 2×3 system has rank two, hence one free variable.

(iii) Show that the 3×2 linear system below is consistent with coefficient matrix having rank two and solution having no free variables,

$$x_1 + 3x_2 = 2,$$
$$2x_1 + 2x_2 = 0,$$
$$3x_1 + x_2 = -2.$$

Solution: Perform the following Gauss operations on the augmented matrix,

$$\left[\begin{array}{cc|c} 1 & 3 & 2 \\ 2 & 2 & 0 \\ 3 & 1 & -2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & 2 \\ 0 & -4 & -4 \\ 0 & -8 & -8 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right].$$

We see that the 3×2 system has rank two, hence no free variables. ◁

1.3.3. **Inconsistent linear systems.** Not every linear system has solutions. We now use both the reduced echelon form of the coefficent matrix and of the augmented matrix to characterize inconsistent linear systems. We start with an example.

Example 1.3.10: Show that the following 2×2 linear system has no solutions,

$$2x_1 - x_2 = 0 \tag{1.16}$$
$$-\frac{1}{2}x_1 + \frac{1}{4}x_1 = -\frac{1}{4}. \tag{1.17}$$

Solution: One way to see that there is no solution is the following: Multiplying the second equation by -4 one obtains the equation

$$2x_1 - x_2 = 1,$$

whose solutions form a parallel line to the line given in Eq. (1.16). Therefore, the system in Eqs. (1.16)-(1.17) has no solution. A second way to see that the system above has no solution is using Gauss operations. The system above has augmented matrix

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ -\frac{1}{2} & \frac{1}{4} & -\frac{1}{4} \end{array}\right] \to \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

The echelon form above corresponds to the linear system

$$2x_1 - x_2 = 0$$
$$0 = 1.$$

The solutions of this second system coincide with the solutions of Eqs. (1.16)-(1.17). Since the second system has no solution, the system in Eqs. (1.16)-(1.17) has no solution. ◁

This example is a particular cases of the following result.

Theorem 1.3.7. *An $m \times n$ linear system with augmented matrix $[\mathsf{A}|\mathsf{b}]$ is inconsistent iff the reduced echelon form of its augmented matrix contains a row of the form $[0, \cdots, 0 \mid 1]$; equivalently $\text{rank}(\mathsf{A}) < \text{rank}(\mathsf{A}|\mathsf{b})$. Furthermore, a consistent system contains:*

(i) A unique solution iff it has no free variables; equivalently $\text{rank}(\mathsf{A}) = n$.

(ii) Infinitely many solutions iff it has at least one free variable; equivalently $\text{rank}(\mathsf{A}) < n$.

This Theorem says that a system is inconsistent iff its augmented matrix has a pivot in the source vector column, which is the last column in an augmented matrix. In that case the system includes an equation of the form $0 = 1$, which has no solution.

The idea of the proof if this Theorem is to study all possible reduced echelon forms $\mathsf{E_A}$ of an arbitrary matrix A, and then to study all possible augmented matrices $[\mathsf{E_A}|\mathsf{c}]$. One then concludes that there are three main cases, no solutions, unique solutions, or infinitely many solutions.

Proof of Theorem 1.3.7: We only give the proof in the case of 3×3 linear systems. The reduced echelon form $\mathsf{E_A}$ of a 3×3 matrix A determines a 3×4 augmented matrix $[\mathsf{E_A}|\mathsf{c}]$. There are 14 possible forms for this matrix. We start with the case of three pivots:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & * \\ 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \end{array}\right], \quad \left[\begin{array}{ccc|c} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 1 \end{array}\right], \quad \left[\begin{array}{ccc|c} 1 & * & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{array}\right], \quad \left[\begin{array}{ccc|c} 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{array}\right].$$

In the first case we have a unique solution, and $\text{rank}(\mathsf{A}) = 3$. The other three cases correspond to no solutions. We now continue with the case of two pivots, which contains six possibilities, with the first three given by

$$\left[\begin{array}{ccc|c} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 1 & * & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 \end{array}\right],$$

which correspond to infinitely many solutions, with one free variable, $\text{rank}(\mathsf{A}) = 2$. The other three possibilities are given by

$$\left[\begin{array}{ccc|c} 1 & * & * & * \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 0 & 1 & * & * \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right];$$

which correspond to no solution. Finally, we have the case of one pivot. It contains four possibilities, the first three of them are given by

$$\left[\begin{array}{ccc|c} 1 & * & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right];$$

which correspond to infinitely many solutions, with two free variables and $\text{rank}(\mathsf{A}) = 1$. The last possibility is the trivial case

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

which has no solutions. This establishes the Theorem for 3×3 linear systems. □

EXAMPLE 1.3.11: Sow that the 2×2 linear system below is inconsistent:

$$2x_1 - x_2 = 1$$
$$-\frac{1}{2}\,x_1 + \frac{1}{4}\,x_1 = -\frac{1}{4}.$$

SOLUTION: The proof of the statement above is the following: Gauss-Jordan operations transform the augmented matrix of the system as follows,

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ -\frac{1}{2} & \frac{1}{4} & -\frac{1}{4} \end{array}\right] \to \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & -\frac{1}{4} \end{array}\right] \to \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

Since there is a line of the form $[0, 0|1]$, the system is inconsistent. We can also say that the coefficient matrix has rank one, but the definition of free variables does not apply, since the system is inconsistent. ◁

EXAMPLE 1.3.12: Find all numbers h and k such that the system below has only one, many, or no solutions,

$$x_1 + hx_2 = 1$$
$$x_1 + 2x_2 = k.$$

SOLUTION: Start finding the associated augmented matrix and reducing it into echelon form,

$$\left[\begin{array}{cc|c} 1 & h & 1 \\ 1 & 2 & k \end{array}\right] \to \left[\begin{array}{cc|c} 1 & h & 1 \\ 0 & 2-h & k-1 \end{array}\right].$$

Suppose $h \neq 2$, for example set $h = 1$, then

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & k-1 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 2-k \\ 0 & 1 & k-1 \end{array}\right],$$

so the system has a unique solution for all values of k. (The same conclusion holds if one sets h to any number different of 2.) Suppose now that $h = 2$, then,

$$\left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & 0 & k-1 \end{array}\right].$$

If $k = 1$ then

$$\left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & 0 & 0 \end{array}\right] \quad \Rightarrow \quad \begin{cases} x_1 = 1 - 2x_2, \\ x_2 : \text{ free variable.} \end{cases}$$

so there are infinitely many solutions. If $k \neq 1$, the system is inconsistent, since

$$\left[\begin{array}{cc|c} 1 & 2 & 1, \\ 0 & 0 & k-1 \neq 0. \end{array}\right]$$

Summarizing, for $h \neq 2$ the system has a unique solution for every k. If $h = 2$ and $k = 1$ the system has infinitely many solutions, if $h = 2$ and $k \neq 1$ the system has no solution. ◁

Further reading. See Sections 2.1, 2.2 and 2.3 in Meyer's book [3] for a detailed discussion on echelon forms and rank, reduced echelon forms and inconsistent systems, respectively. Again, see Section 1.2 in Lay's book [2] for a summary of echelon forms and the Gauss-Jordan method. Section 1.3 in Strang's book [4] also helps.

1.3.4. **Exercises.**

1.3.1.- Find the rank and the pivot columns of the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 4 & 2 & 2 \\ 3 & 6 & 3 & 4 \end{bmatrix}.$$

1.3.2.- Find all the solutions x of the linear system $\mathsf{Ax} = \mathsf{0}$, where the matrix A is given by

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

1.3.3.- Find all the solutions x of the linear system $\mathsf{Ax} = \mathsf{b}$, where the matrix A and the vector b are given by

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & -7 \\ 3 & 2 & 0 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 3 \\ -4 \\ 1 \end{bmatrix}.$$

1.3.4.- Construct a 3×4 matrix A and 3-vectors b, c, such that $[\mathsf{A}|\mathsf{b}]$ is the augmented matrix of a consistent system and $[\mathsf{A}|\mathsf{c}]$ is the augmented matrix of an inconsistent system.

1.3.5.- Let A be an $m \times n$ matrix having $\text{rank}(\mathsf{A}) = m$. Explain why the system with augmented matrix $[\mathsf{A}|\mathsf{b}]$ is consistent for every m-vector b.

1.3.6.- Consider the following system of linear equations, where k represents any real number,

$$2x_1 + kx_2 = 1,$$
$$-4x_1 + 2x_2 = 4.$$

(a) Find all possible values of the number k such that the system above is **in**consistent.

(b) Set $k = 2$. In this case, find the solution $\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

1.4. Non-homogeneous equations

In this Section we introduce the definitions of homogeneous and non-homogeneous linear systems and discuss few relations between their solutions. We start, however, introducing new notation. We first define a vector form for the solution of a linear system and we then introduce a matrix-vector product in order to express a linear system in a compact way. One advantage of this notation appears in this Section: It is simple to realize that solutions of a non-homogeneous linear system are translations of solutions of the associated homogeneous system. A somehow deeper insight on the relation between matrices and vectors can be obtained from the matrix-vector product, which will be discussed in the next Chapter. Here we can summarize this insight as follows: A matrix can be thought as a function acting on the space of vectors.

1.4.1. **Matrix-vector product.** We start generalizing the vector notation to include the unknowns of a linear system. First, recall that given an $m \times n$ linear system

$$A_{11}x_1 + \cdots + A_{1n}x_n = b_1, \tag{1.18}$$

$$\vdots$$

$$A_{m1}x_1 + \cdots + A_{mn}x_n = b_m, \tag{1.19}$$

we have introduced the $m \times n$ coefficients matrix $\mathsf{A} = [A_{ij}]$ and the source m-vector $\mathsf{b} = [b_i]$, where $i = 1, \cdots, m$ and $j = 1, \cdots, n$. Now, introduce the *unknown n-vector* $\mathsf{x} = [x_j]$. The matrix A and the vectors b, x can be used to express a linear system if we introduce an operation between a matrix and a vector. The result of this operation must be the left-hand side in Eqs. (1.18)-(1.19).

Definition 1.4.1. *The **matrix-vector product** of an $m \times n$ matrix $\mathsf{A} = [A_{ij}]$ and an n-vector $\mathsf{x} = [x_j]$, where $i = 1, \cdots, m$ an $j = 1, \cdots, n$, is the m-vector*

$$\mathsf{Ax} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + \cdots + A_{1n}x_n \\ \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n \end{bmatrix}.$$

We now use the matrix, vectors above together with the matrix-vector product to express a linear system in a compact notation.

Theorem 1.4.2. *Given the algebraic linear system in Eqs. (1.18)-(1.19), introduce the coefficient matrix A, the unknown vector x, and the source vector b, as follows,*

$$\mathsf{A} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}, \qquad \mathsf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Then, the algebraic linear system can be written as

$$\mathsf{Ax} = \mathsf{b}.$$

Proof of Theorem 1.4.2: From the definition of the matrix-vector product we have that

$$\mathsf{Ax} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + \cdots + A_{1n}x_n \\ \vdots \\ A_{n1}x_1 + \cdots + A_{1n}x_n \end{bmatrix}.$$

Then, we conclude that

$$\left.\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= b_1, \\ &\vdots \\ A_{n1}x_1 + \cdots + A_{nn}x_n &= b_n, \end{aligned}\right\} \quad \Leftrightarrow \quad \begin{bmatrix} A_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ A_{n1}x_1 + \cdots + A_{1n}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \quad \Leftrightarrow \quad \mathsf{Ax} = \mathsf{b}.$$

□

Therefore, an $m \times n$ linear system with augmented matrix $[\mathsf{A}|\mathsf{b}]$ can now be expressed as using the matrix-vector product as $\mathsf{Ax} = \mathsf{b}$, where x is the variable n-vector and b is the usual source m-vector.

EXAMPLE 1.4.1: Use the matrix-vector product to express the linear system

$$2x_1 - x_2 = 3,$$
$$-x_1 + 2x_2 = 0.$$

SOLUTION: The matrix of coefficient A, the variable vector x and the source vector b are respectively given by

$$\mathsf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

The coefficient matrix can be written as

$$\mathsf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \left[\mathsf{A}_{:1}, \mathsf{A}_{:2}\right], \quad \mathsf{A}_{:1} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \mathsf{A}_{:2} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

The linear system above can be written in the compact way $\mathsf{Ax} = \mathsf{b}$, that is,

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

We now verify that the notation above actually represents the original linear system:

$$\begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} x_1 + \begin{bmatrix} -1 \\ 2 \end{bmatrix} x_2 = \begin{bmatrix} 2x_1 \\ -x_1 \end{bmatrix} + \begin{bmatrix} -x_2 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 \end{bmatrix},$$

which indeed is the original linear system. ◁

EXAMPLE 1.4.2: Use the matrix-vector product to express the 2×3 linear system

$$2x_1 - 2x_2 + 4x_3 = 6,$$
$$x_1 + 3x_2 + 2x_3 = 10.$$

SOLUTION: The matrix of coefficients A, variable vector x and source vector b are given by

$$\mathsf{A} = \begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix}, \quad \mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}.$$

Therefore, the linear system above can be written as $\mathsf{Ax} = \mathsf{b}$, that is,

$$\begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}.$$

We now verify that this notation reproduces the linear system above:

$$\begin{bmatrix} 6 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} -2 \\ 3 \end{bmatrix} x_2 + \begin{bmatrix} 4 \\ 2 \end{bmatrix} x_3 = \begin{bmatrix} 2x_1 - 2x_2 + 4x_3 \\ x_1 + 3x_2 + 2x_3 \end{bmatrix},$$

which indeed is the linear system above. ◁

EXAMPLE 1.4.3: Use the matrix-vector product to express the 3×2 linear system

$$\begin{aligned} x_1 - x_2 &= 0, \\ -x_1 + x_2 &= 2, \\ x_1 + x_2 &= 0. \end{aligned}$$

SOLUTION: Using the matrix-vector product we get

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}.$$

We now verify that the notation above actually represents the original linear system:

$$\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} x_2 = \begin{bmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_1 + x_2 \end{bmatrix},$$

which is indeed the linear system above. ◁

1.4.2. **Linearity of matrix-vector product.** Introduce a column-vector notation for the coefficient matrix, that is,

$$\mathsf{A} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix} = \begin{bmatrix} \mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n} \end{bmatrix},$$

where $\mathsf{A}_{:j}$ is the j-th column of the coefficient matrix A. Using this notation we can rewrite a matrix-vector product as a linear combination of the matrix column vectors $\mathsf{A}_{:j}$. This is our first result.

Theorem 1.4.3. *The matrix-vector product of an $m \times n$ matrix $\mathsf{A} = \begin{bmatrix} \mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n} \end{bmatrix}$ and an n-vector $\mathsf{x} = [x_i]$ can be written as follows,*

$$\mathsf{Ax} = \mathsf{A}_{:1} x_1 + \cdots + \mathsf{A}_{:n} x_n.$$

Proof of Theorem 1.4.3: Write down the matrix-vector product of the $m \times n$ matrix $\mathsf{A} = [A_{ij}]$ and the n-vector $\mathsf{x} = [x_j]$, that is,

$$\mathsf{Ax} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + \cdots + A_{1n}x_n \\ \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n \end{bmatrix} = \begin{bmatrix} A_{11}x_1 \\ \vdots \\ A_{m1}x_1 \end{bmatrix} + \cdots + \begin{bmatrix} A_{1n}x_n \\ \vdots \\ A_{mn}x_n \end{bmatrix}.$$

The expression on the far right can be rewritten as

$$\mathsf{Ax} = \begin{bmatrix} A_{11} \\ \vdots \\ A_{m1} \end{bmatrix} x_1 + \cdots + \begin{bmatrix} A_{1n} \\ \vdots \\ A_{mn} \end{bmatrix} x_n \quad \Rightarrow \quad \mathsf{Ax} = \mathsf{A}_{:1} x_1 + \cdots + \mathsf{A}_{:n} x_n.$$

This establishes the Theorem. □

We are now ready to show that the matrix-vector product has an important property: It preserves the linear combination of vectors. We say that the matrix-vector vector is a *linear operation*. This property is summarized below.

Theorem 1.4.4. *For every $m \times n$ matrix A, every n-vectors x, y and for every numbers a, b, the matrix-vector product satisfies that*

$$\mathsf{A}(a\mathsf{x} + b\mathsf{y}) = a\,\mathsf{A}\mathsf{x} + b\,\mathsf{A}\mathsf{y}.$$

In words, the Theorem says that the matrix-vector product of a linear combination of vectors is the linear combination of the matrix-vector products. The expression above contains the particular cases $a = b = 1$ and $b = 0$, which are respectively given by

$$\mathsf{A}(\mathsf{x} + \mathsf{y}) = \mathsf{A}\mathsf{x} + \mathsf{A}\mathsf{y}, \qquad \mathsf{A}(a\mathsf{x}) = a\,\mathsf{A}\mathsf{x}.$$

Proof of Theorem 1.4.4: From the definition of the matrix-vector product we see that:

$$\mathsf{A}(a\mathsf{x} + b\mathsf{y}) = [\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}] \begin{bmatrix} ax_1 + by_1 \\ \vdots \\ ax_n + by_n \end{bmatrix} = \mathsf{A}_{:1}(ax_1 + by_1) + \cdots + \mathsf{A}_{:n}(ax_n + by_n).$$

Reorder terms in the expression above to get,

$$\mathsf{A}(a\mathsf{x} + b\mathsf{y}) = a\big(\mathsf{A}_{:1}x_1 + \cdots + \mathsf{A}_{:n}x_n\big) + b\big(\mathsf{A}_{:1}y_1 + \cdots + \mathsf{A}_{:n}y_n\big) = a\,\mathsf{A}\mathsf{x} + b\,\mathsf{A}\mathsf{y}.$$

This establishes the Theorem. □

1.4.3. **Homogeneous linear systems.** All possible linear systems can be classified into two main classes, homogeneous and non-homogeneous, depending whether the source vector vanishes or not. This classification will be useful to express solutions of non-homogeneous systems in terms of solutions of the associated homogeneous system.

Definition 1.4.5. *The $m \times n$ linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$ is called **homogeneous** iff it holds that $\mathsf{b} = \mathsf{0}$; and is called **non-homogeneous** iff it holds that $\mathsf{b} \neq \mathsf{0}$.*

Every $m \times n$ homogeneous linear system has at least one solution, given by $\mathsf{x} = \mathsf{0}$, called the *trivial* solution of the homogeneous system. The following example shows that homogeneous linear systems can also have non-trivial solutions.

EXAMPLE 1.4.4: Find all solutions of the system $\mathsf{A}\mathsf{x} = \mathsf{0}$, with coefficient matrix

$$\mathsf{A} = \begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix}.$$

SOLUTION: The linear system above can be written in the usual way as follows,

$$-2x_1 + 4x_2 = 0, \tag{1.20}$$

$$x_1 - 2x_2 = 0. \tag{1.21}$$

The solution of this system can be found performing the Gauss elimination operations

$$\begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix} \to \begin{bmatrix} -2 & 4 \\ 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = 2x_2, \\ x_2 : \text{free variable.} \end{cases}$$

We see above that the coefficient matrix of this system has rank one, so the solutions have one free variable. The set of all solutions of the linear system above is given by

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} x_2, \qquad x_2 \in \mathbb{R}.$$

We conclude that the set of all solutions of this linear system can be identified with the set of points that belong to the line shown in Fig. 14. ◁

It is also possible that an homogeneous linear system has only the trivial solution, like the following example shows.

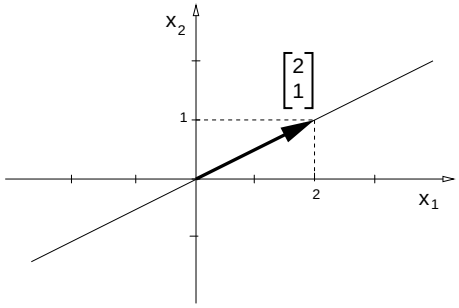

FIGURE 14. We plot the solutions of the homogeneous linear system given in Eqs. (1.20)-(1.21).

EXAMPLE 1.4.5: Find all solutions of the system $\mathsf{Ax} = \mathsf{0}$ with coefficient matrix

$$\mathsf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

SOLUTION: The linear system above can be written in the usual way as follows,

$$2x_1 - x_2 = 0,$$
$$-x_1 + 2x_2 = 0.$$

The solutions of this system can be found performing the Gauss elimination operations

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 2 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = 0, \\ x_2 = 0. \end{cases}$$

We see that the coefficient matrix of this system has rank two, so the solutions have no free variable. The solution is unique and is the trivial solution $\mathsf{x} = \mathsf{0}$. ◁

Examples 1.4.4 and 1.4.5 are particular cases of the following statement: An $m \times n$ homogeneous linear system has non-trivial solutions iff the system has at least one free variable. We show more examples of this statement.

EXAMPLE 1.4.6: Find all solutions of the 2×3 homogeneous linear system $\mathsf{Ax} = \mathsf{0}$ with coefficient matrix

$$\mathsf{A} = \begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix}.$$

SOLUTION: The linear system above can be written in the usual way as follows,

$$2x_1 - 2x_2 + 4x_3 = 0, \tag{1.22}$$
$$x_1 + 3x_2 + 2x_3 = 0. \tag{1.23}$$

The solutions of system above can be obtained through the Gauss elimination operations

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & -2 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 2 \\ 0 & -8 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = -2x_3, \\ x_2 = 0, \\ x_3 : \text{free variable.} \end{cases}$$

The set of all solutions of the linear system above can be written in vector notation,

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_3 \\ 0 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} x_3, \qquad x_3 \in \mathbb{R}.$$

In Fig. 15 we emphasize that the solution vector x belongs to the space $\mathbb{R}^3$, while the column vectors of the coefficient matrix of this same system belong to the space $\mathbb{R}^2$. ◁

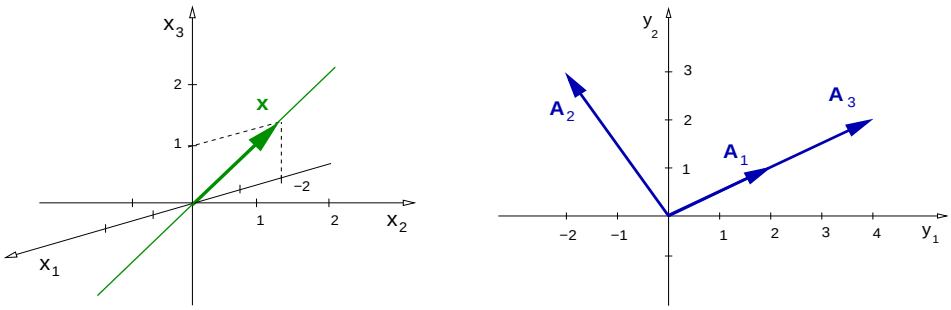

FIGURE 15. The picture on the left represents the solutions of the homogeneous linear system given in Eq. (1.22)-(1.23), which are 3-vectors, elements in $\mathbb{R}^3$. The picture on the right represents the column vectors of the coefficient matrix in this system which are 2-vectors, elements in $\mathbb{R}^2$.

EXAMPLE 1.4.7: Find all solutions to the linear system $\mathsf{Ax} = \mathsf{0}$, with coefficient matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 8 \end{bmatrix}.$$

SOLUTION: We only need to use the Gauss-Jordan method on the coefficient matrix

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 4 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = -3x_2 - 4x_3 \\ x_2,\ x_3 \ \text{: free variables.} \end{cases}$$

In this case the solution can be expressed in vector notation as follows

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3x_2 - 4x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix} x_3.$$

We conclude that all solutions of the homogeneous linear system above are all possible linear combinations of the vectors

$$\mathsf{u}_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathsf{u}_2 = \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}.$$

◁

1.4.4. **The span of vector sets.** From Examples 1.4.4-1.4.7 above we see that solutions of homogeneous linear systems can be expressed as the sets of all possible linear combinations of particular vectors. Since this type of set will appear very often in our studies, it is convenient to give such sets a name. We introduce the span of a finite set of vectors as the infinite set of all possible linear combinations of this finite set of vectors.

Definition 1.4.6. *The **span** of a finite set* $U = \{\mathsf{u}_1, \cdots, \mathsf{u}_k\} \subset \mathbb{R}^n$*, with* $k \geqslant 1$*, denoted as* $\text{Span}(U)$*, is the set given by*

$$\text{Span}(U) = \{\mathsf{x} \in \mathbb{R}^n \ : \ \mathsf{x} = x_1\mathsf{u}_1 + \cdots + x_n\mathsf{u}_n, \quad \forall\ x_1, \cdots, x_n \in \mathbb{R}\}$$

Recall that the symbol "$\forall$" means "for all". Using this definition we express the solutions of $\mathsf{Ax} = \mathsf{0}$ in Example 1.4.7 as

$$\mathsf{x} \in \text{Span}\Big(\Big\{\mathsf{u}_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}\Big\}\Big).$$

In this case, the set of all solutions forms a plane in $\mathbb{R}^3$, which contains the vectors u_1 and u_2. In the case of Example 1.4.4 the solutions x belong to a line in $\mathbb{R}^2$ given by

$$\mathsf{x} \in \text{Span}\Big(\Big\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\Big\}\Big).$$

EXAMPLE 1.4.8: Find the set $S = \text{Span}\Big(\Big\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}\Big\}\Big) \subset \mathbb{R}^2$.

SOLUTION: Since the vectors are not proportional to each other, the set of all linear combinations of these vectors is the whole plane. We conclude that $S = \mathbb{R}^2$. ◁

EXAMPLE 1.4.9: Find the set $S = \text{Span}\Big(\Big\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \end{bmatrix}\Big\}\Big) \subset \mathbb{R}^2$.

SOLUTION: Since the vectors lay on a line, the set of all linear combinations of these vectors also belongs to the same line. We conclude that $S = \Big\{a \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ a \in \mathbb{R}\Big\}$. ◁

1.4.5. **Non-homogeneous linear systems.** Spans of vector sets are not only useful to express solution sets to homogeneous equations, they are also useful to characterize when a non-homogeneous linear system is consistent.

Theorem 1.4.7. *An $m \times n$ linear system $\mathsf{Ax} = \mathsf{b}$, with coefficient matrix $\mathsf{A} = \big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big]$, is consistent iff $\mathsf{b} \in \text{Span}\big(\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\}\big)$.*

In words, a non-homogeneous linear system is consisten iff the source vector belongs to the Span of the coefficient matrix column vectors.

Proof of Theorem 1.4.7:

($\Rightarrow$) Suppose that $\mathsf{x} = [x_i]$ is any solution of the linear system $\mathsf{Ax} = \mathsf{b}$. Using the column vector notation for matrix A we get that

$$\mathsf{b} = \mathsf{Ax} = \mathsf{A}_{:1}x_1 + \cdots + \mathsf{A}_{:n}x_n.$$

This last expression says that $\mathsf{b} \in \text{Span}\big(\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\}\big)$.

($\Leftarrow$) If $\mathsf{b} \in \text{Span}\big(\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\}\big)$, then there exists constants $c_1, \cdots, c_n$ such that

$$\mathsf{b} = \mathsf{A}_{:1}c_1 + \cdots + \mathsf{A}_{:n}c_n = \mathsf{Ac}, \qquad \mathsf{c} = [c_i].$$

This last equation says that the vector $\mathsf{x} = \mathsf{c}$ is a solution of the linear system $\mathsf{Ax} = \mathsf{b}$, so the system is consistent. This establishes the Theorem. □

Knowing the solutions of an homogeneous linear system provides important information about the solutions of an inhomogeneous linear system with the same coefficient matrix. The next result establishes this relation in a precise way.

Theorem 1.4.8. *If the $m \times n$ linear system $\mathsf{Ax} = \mathsf{b}$ is consistent and the vector x_p is one particular solution of this linear system, then any solution x to this system can be decomposed as $\mathsf{x} = \mathsf{x}_p + \mathsf{x}_h$, where the vector x_h is a solution of the homogeneous linear system $\mathsf{Ax}_h = \mathsf{0}$.*

Proof of Theorem 1.4.8: We know that the vector x_p is a solution of the linear system, that is, $\mathsf{A}\mathsf{x}_p = \mathsf{b}$. Suppose that there is any other solution x of the same system, $\mathsf{A}\mathsf{x} = \mathsf{b}$. Then, their difference $\mathsf{x}_h = \mathsf{x} - \mathsf{x}_p$ satisfies the homogeneous equation $\mathsf{A}\mathsf{x}_h = \mathsf{0}$, since

$$\mathsf{A}\mathsf{x}_h = \mathsf{A}(\mathsf{x} - \mathsf{x}_p) = \mathsf{A}\mathsf{x} - \mathsf{A}\mathsf{x}_p = \mathsf{b} - \mathsf{b} = \mathsf{0},$$

where the linearity of the matrix-vector product, proven in Theorem 1.4.4, is used in the step $\mathsf{A}(\mathsf{x} - \mathsf{x}_p) = \mathsf{A}\mathsf{x} - \mathsf{A}\mathsf{x}_p$. We have shown that $\mathsf{x}_h = \mathsf{x} - \mathsf{x}_p$, that is, any solution x of the non-homogeneous system can be written as $\mathsf{x} = \mathsf{x}_p + \mathsf{x}_h$. This establishes the Theorem. □

We say that the solution to a non-homogeneous linear system is written in *vector form* or in *parametric form* when it is expressed as in Theorem 1.4.8, that is, as $\mathsf{x} = \mathsf{x}_p + \mathsf{x}_h$, where the vector x_h is a solution of the homogeneous linear system, and the vector x_p is any solution of the non-homogeneous linear system.

EXAMPLE 1.4.10: Find all solutions of the linear system and write them in parametric form,

$$\begin{bmatrix} 2 & -4 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}. \tag{1.24}$$

SOLUTION: We first find the solutions of this non-homogeneous linear system using Gauss elimination operations,

$$\left[\begin{array}{cc|c} 1 & -2 & 3 \\ 2 & -4 & 6 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & -2 & 3 \\ 0 & 0 & 0 \end{array}\right] \quad \Rightarrow \quad \begin{cases} x_1 = 2x_2 + 3, \\ x_2 : \text{free variable.} \end{cases}$$

Therefore, the set of all solutions of the linear system above is given by

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_2 + 3 \\ x_2 \end{bmatrix} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

In this case we see that

$$\mathsf{x}_p = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{(by choosing } x_2 = 0\text{), and} \quad \mathsf{x}_h = \begin{bmatrix} 2 \\ 1 \end{bmatrix} x_2.$$

The vector x_p is the particular solution to the non-homogeneous system given by $x_2 = 0$, while it is not difficult to check that x_h above is solution of the homogeneous equation $\mathsf{A}\mathsf{x}_h = \mathsf{0}$. In Fig. 16 we plot these solutions on the plane. The solution of the non-homogeneous system is the translation by x_p of the solutions of the homogeneous system. ◁

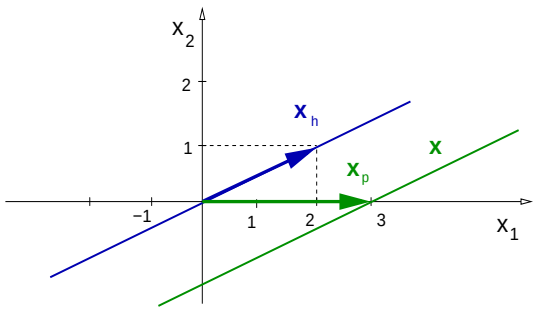

FIGURE 16. The blue line represents solutions to the homogeneous system in Eq. (1.24). The green line represents the solutions to the non-homogeneous system in Eq. (1.24), which is the translation by x_p of the line passing by the origin.

Further reading. See Sections 2.4 and 2.5 in Meyer's book [3] for a detailed discussion on homogeneous and non-homogeneous linear systems, respectively. See Section 1.4 in Lay's book [2] for a detailed discussion of the matrix-vector product, and Section 1.5 for detailed discussions on homogeneous and non-homogeneous linear systems.

1.4.6. **Exercises.**

1.4.1.- Find the general solution of the homogeneous linear system

$$\begin{aligned} x_1 + 2x_2 + x_3 + 2x_4 &= 0 \\ 2x_1 + 4x_2 + x_3 + 3x_4 &= 0 \\ 3x_1 + 6x_2 + x_3 + 4x_4 &= 0. \end{aligned}$$

1.4.2.- Find all the solutions x of the linear system $\mathsf{Ax} = \mathsf{b}$, where A and b are given by

$$\mathsf{A} = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 1 & 8 \\ 1 & -1 & 1 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},$$

and write these solutions in parametric form, that is, in terms of column vectors.

1.4.3.- Prove the following statement: If the vectors c and d are solutions of the homogeneous linear system $\mathsf{Ax} = \mathsf{0}$, then $\mathsf{c} + \mathsf{d}$ is also a solution.

1.4.4.- Find the general solution of the nonhomogeneous linear system

$$\begin{aligned} x_1 + 2x_2 + x_3 + 2x_4 &= 3 \\ 2x_1 + 4x_2 + x_3 + 3x_4 &= 4 \\ 3x_1 + 6x_2 + x_3 + 4x_4 &= 5. \end{aligned}$$

1.4.5.- Suppose that the solution to a system of linear equation is given by

$$\begin{aligned} x_1 &= 5 + 4x_3 \\ x_2 &= -2 - 7x_3 \\ x_3 &\quad \text{free.} \end{aligned}$$

Use column vectors to describe this set as a line in $\mathbb{R}^3$.

1.4.6.- Suppose that the solution to a system of linear equation is given by

$$\begin{aligned} x_1 &= 3x_4 \\ x_2 &= 8 + 4x_4 \\ x_3 &= 2 - 5x_4 \\ x_4 &\quad \text{free.} \end{aligned}$$

Use column vectors to describe this set as a line in $\mathbb{R}^4$.

1.4.7.- Consider the following system of linear equations, where k represents any real number,

$$\begin{bmatrix} 2 & 2 & 3 \\ 4 & 8 & 12 \\ 6 & 2 & k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \\ 4 \end{bmatrix}.$$

(a) Find all possible values of the number k such that the system above has a unique solution.

(b) Find all possible values of the number k such that the system above has infinitely many solutions, and express those solutions in parametric form.

1.5. Floating-point numbers

1.5.1. **Main definitions.** Floating-point numbers are a finite subset of the rational numbers. Many different types of floating-point numbers exist, all of them are characterized by having a finite number of digits when written in a particular base. Digital computers use floating-point numbers to carry out almost every arithmetic operation. When an $m \times n$ algebraic linear system is solved using a computer, every Gauss operations is performed in a particular set of floating-point numbers. In this Section we study what type of approximations occur in this process.

Definition 1.5.1. *A non-zero rational number x is a **floating-point number** in base $b \in \mathbb{N}$, of precision $p \in \mathbb{N}$, with exponent $n \in \mathbb{Z}$ in the range $-N \leqslant n \leqslant N \in \mathbb{N}$, iff there exist integers d_i, for $i = 1, \cdots, p$, satisfying $0 \leqslant d_i \leqslant b-1$ and $d_1 \neq 0$, such that the number x has the form*

$$x = \pm 0.d_1 \cdots d_p \times b^n. \tag{1.25}$$

*We call p the **precision**, b the **base** and N the **exponent range** of the floating point number x. We denote by $\mathbb{F}_{p,b,N}$ the set of all floating-point numbers of fixed precision p, base b and exponent range N.*

In this notes we always work in base $b = 10$. Computers usually work with base $b = 2$, but also with base $b = 16$, and they present their results with base $b = 10$.

Example 1.5.1: The following numbers belong to the set $\mathbb{F}_{3,10,3}$,

$$210 = 0.210 \times 10^3, \quad 1 = 0.100 \times 10, \quad -0.02 = -0.200 \times 10^{-1}, \quad \frac{215}{10^6} = 0.215 \times 10^{-3}.$$

The set $\mathbb{F}_{3,10,3}$ is a finite subset of the rational numbers. The biggest number and the smallest number in absolute value are the following, respectively,

$$0.999 \times 10^3 = 999 \qquad 0.100 \times 10^{-3} = 0.0001.$$

Any number bigger 999 or closer to 0 than 0.0001 does not belong to $\mathbb{F}_{3,10,3}$. Here are other examples of numbers that do not belong to $\mathbb{F}_{3,10,3}$,

$$1000 = 0.100 \times 10^4 \quad 210.3 = 0.2103 \times 10^3, \quad 1.001 = 0.1001 \times 10, \quad 0.000027 = 0.270 \times 10^{-4}.$$

In the first case the number is too big, we need an exponent $n = 4$ to specify it. In the second and third cases we need a precision $p = 4$ to specify those numbers. In the last case the number is too close to zero, we need an exponent $n = -4$ to specify it. ◁

Example 1.5.2: The set of floating-point numbers $\mathbb{F}_{1,10,1}$ is small enough to picture it on the real line. The set of all positive elements in $\mathbb{F}_{1,10,1}$ is shown on Fig. 17, and this is the union of the following three sets,

$$\{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09\} = \{0.i \times 10^{-1}\}_{i=1}^{9};$$

$$\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\} = \{0.i \times 10^{0}\}_{i=1}^{9};$$

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9\} = \{0.i \times 10^{1}\}_{i=1}^{9}.$$

One can see in this example that the elements on $\mathbb{F}_{1,10,1}$ are not homogeneously distributed on the interval $[0, 10]$. The irregular distribution of the floating-point numbers plays an important role when one computes the addition of a small number to a big number. ◁

Example 1.5.3: In Table 1 we show the main set of floating-point numbers used nowadays in computers. For example, the format called Binary64 represents all floating-point numbers

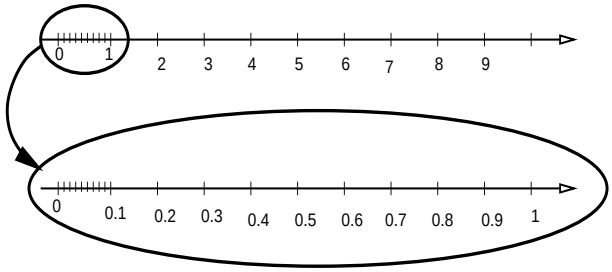

FIGURE 17. All the positive elements in $\mathbb{F}_{1,10,1}$.

in the set $\mathbb{F}_{53,2,1024}$. One of the bigger numbers in this set is 2^{1023}. To have an idea how big is this number, let us rewrite it as 10^x, that is,

$$2^{1023} = 10^x \quad \Leftrightarrow \quad 1023\ln(2) = x\ln(10) \quad \Leftrightarrow \quad x = 307.95...$$

So the biggest number in $\mathbb{F}_{53,2,1024}$ is close to 10^{308}. ◁

Format name	Base b	Digits p	Max. exp. N
Binary 32	2	24	128
Binary 64	2	53	1024
Binary 128	2	113	16384
Decimal l64	10	16	385
Decimal 128	10	34	6145

TABLE 1. List of the parameters b, p and N that determine the floating-point sets $\mathbb{F}_{p,b,N}$, which are most used in computers. The first column presents a standard name given in the scientific computing community to these floating-point formats.

We say that a set $A \subset \mathbb{R}$ is *closed under addition* iff for every two elements in A the sum of these two numbers also belongs to A. The definition of a set $A \subset \mathbb{R}$ being closed under multiplication is similar. The sets of floating-point numbers $\mathbb{F}_{p,b,N} \subset \mathbb{R}$ are not closed under addition or multiplication. This means that the sum of two numbers in $\mathbb{F}_{p,b,N}$ might not belong to $\mathbb{F}_{p,b,N}$. And the multiplication of two numbers in $\mathbb{F}_{p,b,N}$ might not belong to $\mathbb{F}_{p,b,N}$. Here are some examples.

EXAMPLE 1.5.4: Consider the set $\mathbb{F}_{2,10,2}$. It is not difficult to see that $\mathbb{F}_{2,10,2}$ is not closed under multiplication, as the first line below shows. The rest of the example below shows that the sum of two numbers in $\mathbb{F}_{2,10,2}$ does not belong to that set.

$$x = 10^{-3} = 0.10 \times 10^{-2} \in \mathbb{F}_{2,10,2} \quad \Rightarrow \quad x^2 = 10^{-6} = 0.0001 \times 10^{-2} \notin \mathbb{F}_{2,10,2},$$

$$\left.\begin{array}{l} x = 10^{-3} = 0.10 \times 10^{-2} \in \mathbb{F}_{2,10,2}, \\ y = 1 = 0.10 \times 10^{1} \in \mathbb{F}_{2,10,2}, \end{array}\right\} \quad \Rightarrow \quad x + y = 0.001 + 1,$$

$$x + y = 1.001 \quad \Rightarrow \quad x + y = 0.1001 \times 10 \notin \mathbb{F}_{2,10,2}.$$

◁

1.5.2. **The rounding function.** Since the set $\mathbb{F}_{p,b,N}$ is not closed under addition or multiplication, not every arithmetic calculation involving real numbers can be performed in $\mathbb{F}_{p.,b,N}$. A way to perform a sequence of arithmetic operations in $\mathbb{F}_{p,b,N}$ is first, to project the real numbers into the floating-point numbers, and then to perform the calculation. Since the result might not be in the set $\mathbb{F}_{p,b,N}$, one must project again the result onto $\mathbb{F}_{p,b,N}$. The action to project a real number into the floating-point set is called to *round-off* the real number. There are many different ways to do this. We now present a common round-off function.

Definition 1.5.2. *Given the floating-point number set $\mathbb{F}_{p,b,N}$, let $x_N = 0.9 \cdots 9 \times b^N$ be the biggest number in $\mathbb{F}_{p,b,N}$. The* ***rounding function*** *$f_\ell : \mathbb{R} \cap [-x_N, x_N] \to \mathbb{F}_{p,b,N}$ is defined as follow: Given $x \in \mathbb{R} \cap [-x_N, x_N]$, with $x = \pm 0.d_1 \cdots d_p d_{p+1} \cdots \times b^n$ and $-N \leqslant n \leqslant N$, holds*

$$f_\ell(x) = \begin{cases} \pm\, 0.d_1 \cdots d_p \times b^n & \text{if} \quad d_{p+1} < 5, \\ \pm\, (0.d_1 \cdots d_p + b^{-p}) \times b^n & \text{if} \quad d_{p+1} \geqslant 5. \end{cases}$$

EXAMPLE 1.5.5: We now present few numbers not in $\mathbb{F}_{3,10,3}$ and their respective round-offs

$$\begin{aligned} x &= 0.2103 \times 10^3, & f_\ell(x) &= 0.210 \times 10^3, \\ x &= 0.21037 \times 10^3, & f_\ell(x) &= 0.210 \times 10^3, \\ x &= 0.2105 \times 10^3, & f_\ell(x) &= 0.211 \times 10^3, \\ x &= 0.21051 \times 10^3, & f_\ell(x) &= 0.211 \times 10^3. \end{aligned}$$

◁

The rounding function has the following properties:

Proposition 1.5.3. *Given $\mathbb{F}_{p,b,N}$ there always exist x, $y \in \mathbb{R}$ such that*

$$f_\ell\big[f_\ell(x) + f_\ell(y)\big] \neq f_\ell(x + y), \qquad f_\ell\big[f_\ell(x) f_\ell(y)\big] \neq f_\ell(xy).$$

We do not prove this Proposition, we only provide a particular example, in the case that the floating-point number is $\mathbb{F}_{2,10,2}$.

EXAMPLE 1.5.6: The real numbers $x = 21/2$ and $y = 11/2$ add up to $x + y = 16$. Only one of them belongs to $\mathbb{F}_{2,10,2}$, since

$$\begin{aligned} x = 0.105 \times 10^2 \notin \mathbb{F}_{2,10,2} \quad &\Rightarrow \quad f_\ell(x) = 0.11 \times 10^2, \\ y = 0.55 \times 10 \in \mathbb{F}_{2,10,2} \quad &\Rightarrow \quad f_\ell(y) = 0.55 \times 10 = y. \end{aligned}$$

We now verify that for these numbers holds that $f_\ell\big[f_\ell(x) + f_\ell(y)\big] \neq f_\ell(x + y)$, since

$$x + y = 0.16 \times 10^2 \in \mathbb{F}_{2,10,2} \quad \Rightarrow \quad f_\ell(x + y) = 0.16 \times 10^2 = x + y,$$

$$f_\ell\big[f_\ell(x) + f_\ell(y)\big] = f_\ell(0.11 \times 10^2 + 0.55 \times 10) = f_\ell(0.165 \times 10^2) = 0.17 \times 10^2.$$

Therefore, we conclude that

$$17 = f_\ell\big[f_\ell(x) + f_\ell(y)\big] \neq f_\ell(x + y) = 16.$$

◁

1.5.3. **Solving linear systems.** Arithmetic operations are, in general, not possible in $\mathbb{F}_{p.b.N}$, since this set is not closed under these operations. Rounding after each arithmetic operation is a procedure that determines numbers in $\mathbb{F}_{p,b,N}$ which are close to the result of the arithmetic operations. The difference between these two numbers is the error of the procedure, is the error of making calculations in a finite set of numbers. One could think that the Gauss-Jordan method could be carried out in the set $\mathbb{F}_{p,b,N}$, if one round-off each intermediate calculation. This Subsection presents an example that this is not the case. The Gauss-Jordan method is, in general, *not possible* in any set $\mathbb{F}_{p,b,N}$ using rounding after each arithmetic operation.

EXAMPLE 1.5.7: Use the floating-point set $\mathbb{F}_{3,10,3}$ and Gauss operations to find the solution of the 2×2 linear system

$$5\,x_1 + x_2 = 6,$$
$$9.43\,x_1 + 1.57\,x_2 = 11.$$

SOLUTION: Note that the solution is $x_1 = 1$, $x_2 = 1$. The calculation we do now shows that the Gauss-Jordan method is not possible in the set $\mathbb{F}_{3,10,3}$ using round-off functions; the Gauss-Jordan method does not work in this example. The fist step in the Gauss-Jordan method is to compute the augmented matrix of the system and perform a Gauss operation to make the coefficient in the position of a_{21} vanish, that is,

$$\left[\begin{array}{cc|c} 5 & 1 & 6 \\ 9.43 & 1.57 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 5 & 1 & 6 \\ 0 & -0.316 & -0.316 \end{array}\right].$$

This operation cannot be performed in the set $\mathbb{F}_{3,10,3}$ using rounding. In order to understand this, let us review what we did in the calculation above. We multiplied the first row by $-9.43/5$ and add that result to the second row. The result using real numbers is that the new coefficient obtained after this calculation is $\tilde{a}_{21} = 0$. If we do this calculation in the set $\mathbb{F}_{3,10,3}$ using rounding, we have to do the following calculation:

$$\tilde{a}_{21} = f_\ell\Big(9.43 - f_\ell\Big[f_\ell(5)\, f_\ell\Big(\frac{9.43}{5}\Big)\Big]\Big),$$

that is, we round the quotient $-9.43/5$, then we multiply by 5, we round again, then we subtract that from 9.43, and we finally round the result:

$$\begin{aligned} \tilde{a}_{21} &= f_\ell\big(9.43 - f_\ell\big[5\, f_\ell(1.886)\big]\big) \\ &= f_\ell\big(9.43 - f_\ell\big[5\,(1.89)\big]\big) \\ &= f_\ell\big[9.43 - f_\ell(9.45)\big] \\ &= f_\ell\big[9.43 - 9.45\big] \\ &= f_\ell(-0.02) \\ &= -0.02. \end{aligned}$$

Therefore, with this Gauss operation on $\mathbb{F}_{3,10,3}$ using rounding one obtains $\tilde{a}_{21} = -0.02 \neq 0$. The same type of calculation on the other coefficients $\tilde{a}_{22}$, $\tilde{b}_2$, produces the following new augmented matrix

$$\left[\begin{array}{cc|c} 5 & 1 & 6 \\ 9.43 & 1.57 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 5 & 1 & 6 \\ -0.02 & -0.32 & -0.3 \end{array}\right].$$

The Gauss-Jordan method cannot follow unless the coefficient $\tilde{a}_{21} = 0$. But this is not possible in our example. A usual procedure used in scientific computation is to modify the Gauss-Jordan method. The modification introduces further approximation errors in

the calculation. The modification in our example is the following: Replace the coefficient $\tilde{a}_{21} = -0.02$ by $\tilde{a}_{21} = 0$. The modified Gauss-Jordan method in our example is given by

$$\left[\begin{array}{cc|c} 5 & 1 & 6 \\ 9.43 & 1.57 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 5 & 1 & 6 \\ 0 & -0.32 & -0.3 \end{array}\right]. \tag{1.26}$$

What we have done here is not a rounding error. It is a modification of the Gauss-Jordan method to find an approximate solution of the linear system in the set $\mathbb{F}_{3,10,3}$. The rest of the calculation to find the solution of the linear system is the following:

$$\left[\begin{array}{cc|c} 5 & 1 & 6 \\ 0 & -0.32 & -0.3 \end{array}\right] \to \left[\begin{array}{cc|c} 5 & 1 & 6 \\ 0 & 1 & f_\ell\Big(\frac{0.3}{0.32}\Big) \end{array}\right];$$

since

$$f_\ell\Big(\frac{0.3}{0.32}\Big) = f_\ell(0.9375) = 0.938,$$

we have that

$$\left[\begin{array}{cc|c} 5 & 1 & 6 \\ 0 & 1 & f_\ell\Big(\frac{0.3}{0.32}\Big) \end{array}\right] \to \left[\begin{array}{cc|c} 5 & 1 & 6 \\ 0 & 1 & 0.938 \end{array}\right] \to \left[\begin{array}{cc|c} 5 & 0 & f_\ell(6 - 0.938) \\ 0 & 1 & 0.938 \end{array}\right];$$

since

$$f)\ell(6 - 0.938) = f_\ell(5.062) = 5.06,$$

we also have that

$$\left[\begin{array}{cc|c} 5 & 0 & 5.06 \\ 0 & 1 & 0.938 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & f_\ell\Big(\frac{5.06}{5}\Big) \\ 0 & 1 & 0.938 \end{array}\right];$$

since

$$f_\ell\Big(\frac{5.06}{5}\Big) = f_\ell(1.012) = 1.01,$$

we conclude that

$$\left[\begin{array}{cc|c} 1 & 0 & 1.01 \\ 0 & 1 & 0.938 \end{array}\right] \quad \Rightarrow \quad \begin{array}{l} x_1 = 0.101 \times 10, \\ x_2 = 0.938. \end{array}.$$

We conclude that the solution in the set $\mathbb{F}_{3,10,3}$ differs from the exact solution $x_1 = 1$, $x_2 = 1$. The errors in the result are produced by rounding errors and by the modification of the Gauss-Jordan method discussed in Eq. (1.26). ◁

We finally comment that the round-off error becomes important when adding a small number to a big number, or when dividing by a small number. Here is an example of the former case.

Example 1.5.8: Add together the numbers $x = 10^3$ and $y = 4$ in the set $\mathbb{F}_{3,10,4}$.

Solution: Since $x = 0.100 \times 10^4$ and $y = 0.400 \times 10$, both numbers belong to $\mathbb{F}_{3,10,4}$ and so $f_\ell(x) = x$, $f_\ell(y) = y$. Therefore, their addition is the following,

$$f_\ell(x + y) = f_\ell(1000 + 4) = f_\ell(0.1004 \times 10^3) = 1 \times 10^3 = x.$$

That is, $f(x + y) = x$, and the information of y is completely lost. ◁

1.5.4. **Reducing rounding errors.** We have seen that solving an algebraic linear system in a floating-point number set $\mathbb{F}_{p,b,N}$ introduces rounding errors in the solution. There are several techniques that help keep these errors from becoming an important part of the solution. We comment here on four of these techniques. We present these techniques without any proof.

The first two techniques are implemented before one start to solve the linear system. They are called column scaling and row scaling. The *column scaling* consists in multiplying by a same scale factor a whole column in the linear system. One performs a column scaling when one column of a linear system has coefficients that are far bigger or far smaller than the rest of the matrix. The factor in the column scaling is chosen in order that all the coefficients in the linear system are similar in size. One can interpret the column scaling on the column i of a linear system as changing the physical units of the unknown x_i. The *row scaling* consists in multiplying by the same scale factor a whole equation of the system. One performs a row scaling when one row of the linear system has coefficients that are far bigger or far smaller than the rest of the system. Like in the column scaling, one chooses the scaling factor in order that all the coefficients of the linear system are similar in size.

The last two techniques refer to what sequence of Gauss operations is more efficient in reducing rounding errors. It is well-known that there are many different ways to solve a linear system using Gauss operations in the set of the real numbers $\mathbb{R}$. For example, we now solve the linear system below using two different sequences of Gauss operations:

$$\left[\begin{array}{cc|c} 2 & 4 & 10 \\ 1 & 3 & 7 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 1 & 3 & 7 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right],$$

$$\left[\begin{array}{cc|c} 2 & 4 & 10 \\ 1 & 3 & 7 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & 7 \\ 2 & 4 & 10 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & -2 & -4 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

The solution obtained is independent of the sequences of Gauss operations used to find them. This property does not hold in the floating-point number set $\mathbb{F}_{p,b,N}$. Two different sequences of Gauss operations on the same augmented matrix might produce different approximate solutions when they are performed in $\mathbb{F}_{p,b,N}$. The main idea behind the last two techniques we now present to solve linear systems using floating-point numbers is the following: To find the sequence of Gauss operations that minimize the rounding errors in the approximate solution. The last two techniques are called partial pivoting and complete pivoting.

The *partial pivoting* is the row interchange in a matrix in order to have the biggest coefficient in that column as pivot. Here below we do a partial pivoting for the pivot on the first column:

$$\begin{bmatrix} 10^{-2} & 2 & 1 \\ 1 & 10^3 & 10^2 \\ 10 & 4 & 1 \end{bmatrix} \to \begin{bmatrix} 10 & 4 & 1 \\ 1 & 10^3 & 10^2 \\ 10^{-2} & 2 & 1 \end{bmatrix},$$

that is, use a pivot for the first column the coefficient with 10 instead of the coefficient with 10^{-2} or the coefficient with 1. Now proceed in the usual way:

$$\begin{bmatrix} 10 & 4 & 1 \\ 1 & 10^3 & 10^2 \\ 10^{-2} & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0.4 & 0.1 \\ 1 & 10^3 & 10^2 \\ 10^{-2} & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0.4 & 0.1 \\ 0 & 999.6 & 99.9 \\ 0 & 1.996 & 0.999 \end{bmatrix}.$$

The next step is again to use as pivot the biggest coefficient in the second column. In this case it is the coefficient 999.6, so no further row interchanges are necessary. Repeat this procedure till the last column.

The *complete pivoting* is the row and column interchange in a matrix in order to have as pivot the biggest coefficient in the lower-right block from the pivot position. Here below we

do a complete pivoting for the pivot on the first column:

$$\begin{bmatrix} 10^{-2} & 2 & 1 \\ 1 & 10^3 & 10^2 \\ 10 & 4 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 10^3 & 10^2 \\ 10^{-2} & 2 & 1 \\ 10 & 4 & 1 \end{bmatrix} \to \begin{bmatrix} 10^3 & 1 & 10^2 \\ 2 & 10^{-2} & 1 \\ 4 & 10 & 1 \end{bmatrix}.$$

For the first pivot, the lower-right part of the matrix is the whole matrix. In the example above we used as pivot for the first column the coefficient 10^3 in position $(2, 2)$. We needed to do a a row interchange and then a column interchange. Now proceed in the usual way:

$$\begin{bmatrix} 10^3 & 1 & 10^2 \\ 2 & 10^{-2} & 1 \\ 4 & 10 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 10^{-3} & 10^{-1} \\ 2 & 10^{-2} & 1 \\ 4 & 10 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0.001 & 0.1 \\ 0 & 0.008 & 0.8 \\ 0 & 9.996 & 0.6 \end{bmatrix}.$$

The next step in complete pivoting is to choose the coefficient for the pivot position $(2, 2)$. We have to look in the lower-right block from the pivot position, that is, in the block

$$\begin{bmatrix} 0.008 & 0.8 \\ 9.996 & 0.6 \end{bmatrix}.$$

The biggest coefficient in that block is 9.996, so we have to do a row interchange but we do not need to do column interchanges, that is,

$$\begin{bmatrix} 1 & 0.001 & 0.1 \\ 0 & 0.008 & 0.8 \\ 0 & 9.996 & 0.6 \end{bmatrix} \to \begin{bmatrix} 1 & 0.001 & 0.1 \\ 0 & 9.996 & 0.6 \\ 0 & 0.008 & 0.8 \end{bmatrix}.$$

Repeat this procedure till the last column.

Further reading. See Section 1.5 in in Meyer's book [3] for a detailed discussion on solving linear systems using floating point numbers.

1.5.5. **Exercises.**

1.5.1.- Consider the following system:

$$10^{-3}\,x_1 - x_2 = 1,$$
$$x_1 + x_2 = 0.$$

(a) Solve this system in $\mathbb{F}_{3,10,6}$ with rounding, but without partial or complete pivoting.
(b) Find the system that is exactly satisfied by your solution in (a), and note how close is this system to the original system.
(c) Use partial pivoting to solve this system in $\mathbb{F}_{3,10,5}$ with rounding.
(d) Find the system that is exactly satisfied by your solution in (c), and note how close is this system to the original system.
(e) Solve this system in $\mathbb{R}$ without partial or complete pivoting, and compare this exact solution with the solutions in (a) and (c).
(f) Round the exact solution up to three digits, an compare it with the results from (a) and (c).

1.5.2.- Consider the following system:

$$x_1 + x_2 = 3,$$
$$-10\,x_1 + 10^5\,x_2 = 10^5.$$

(a) Solve this system in $\mathbb{F}_{4,10,6}$ with partial pivoting but no scaling.
(b) Solve this system in $\mathbb{F}_{4,10,6}$ with complete pivoting but no scaling.
(c) Use partial pivoting to solve this system in $\mathbb{F}_{3,10,5}$ with rounding.
(d) This time row scale the original system, and then solve it in $\mathbb{F}_{4,10,6}$ with partial pivoting.
(e) Solve this system in $\mathbb{R}$ and compare this exact solution with the solutions in (a)-(d).

1.5.3.- Consider the linear system

$$-3\,x_1 + x_2 = -2,$$
$$10\,x_1 - 3\,x_2 = 7.$$

Solve this system in $\mathbb{F}_{3,10,6}$ without partial pivoting, and then solve it again with partial pivoting. Compare your results with the exact solution.

CHAPTER 2. MATRIX ALGEBRA

2.1. LINEAR TRANSFORMATIONS

In Sect. 1.4 we introduced the matrix-vector product because it provided a convenient notation to express a system of linear equations in a compact way. Here we study a deeper meaning of the matrix-vector product. It is the key to interpret a matrix as a function acting on vectors. We will see that a matrix is a particular type of function, called linear function or linear transformation. We also show examples of functions determined by matrices.

From now on we consider both real-valued and complex-valued matrices and vectors. We use the notation $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ to mean that $\mathbb{F}$ can be either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Elements in $\mathbb{F}$ are called *scalars*. So a scalar is a real number or a complex number. We denote by $\mathbb{F}^n$ the set of all n-vectors $\mathsf{x} = [x_i]$ with components $x_i \in \mathbb{F}$, where $i = 1, \cdots, n$. Finally we denote by $\mathbb{F}^{m,n}$ the set of all $m \times n$ matrices $\mathsf{A} = [A_{ij}]$ with components $A_{ij} \in \mathbb{F}$, where $i = 1 \cdots m$ and $j = 1, \cdots, n$.

2.1.1. **A matrix is a function.** The matrix-vector product provides a new interpretation for a matrix. A matrix is not only an artifact for a compact notation, it defines a function on the set of vectors. Here is a precise definition.

Definition 2.1.1. *Every $m \times n$ matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ defines the function $\mathsf{A} : \mathbb{F}^n \to \mathbb{F}^m$, where the image of an n-vector x is the m-vector $\mathsf{y} = \mathsf{A}\mathsf{x}$, the matrix-vector product of A and x.*

EXAMPLE 2.1.1: Compute the function defined by the 2×3 matrix $\mathsf{A} = \begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix}$.

SOLUTION: The matrix A defines a function $\mathsf{A} : \mathbb{R}^3 \to \mathbb{R}^2$, since for every $\mathsf{x} \in \mathbb{R}^3$ holds,

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \mathsf{A}\mathsf{x} = \begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{y} = \mathsf{A}\mathsf{x} = \begin{bmatrix} 2x_1 - 2x_2 + 4x_3 \\ x_1 + 3x_2 + 2x_3 \end{bmatrix} \in \mathbb{R}^2.$$

For example, given $\mathsf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{R}^3$, then $\mathsf{y} = \mathsf{A}\mathsf{x} = \begin{bmatrix} 2 & -2 & 4 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} \in \mathbb{R}^2$. ◁

EXAMPLE 2.1.2: Describe the function defined by the 2×2 matrix $\mathsf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

SOLUTION: The matrix A defines a function $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$, which can be interpreted as a *reflection along the horizontal line*. Indeed, the action of the matrix A on an arbitrary element in $\mathsf{x} \in \mathbb{R}^2$ is the following,

$$\mathsf{A}\mathsf{x} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}.$$

Here are particular cases, represented in Fig. 18,

$$\mathsf{A} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \mathsf{A} \begin{bmatrix} -1 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \quad \mathsf{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.1}$$

◁

EXAMPLE 2.1.3: Describe the function defined by the 2×2 matrix $\mathsf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

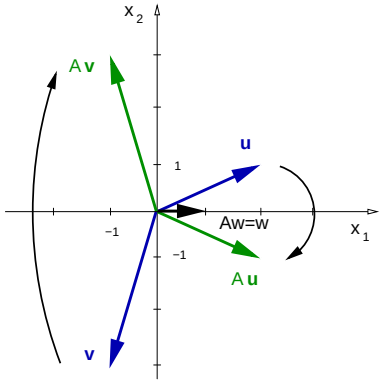

FIGURE 18. We sketch the action of matrix A in Example 2.1.2 on the vectors given in Eq. (2.1), which we called u, v, and w, respectively. Since $A : \mathbb{R}^2 \to \mathbb{R}^2$, we used the same plane to plot both u and Au.

SOLUTION: The matrix A defines a function $A : \mathbb{R}^2 \to \mathbb{R}^2$, which can be interpreted as a *reflection along the line* $x_1 = x_2$, see Fig. 19. Indeed, the action of the matrix A on an arbitrary element in $x \in \mathbb{R}^2$ is the following,

$$Ax = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}.$$

Here are particular cases, represented in Fig. 19,

$$A \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad A \begin{bmatrix} -3 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \end{bmatrix}, \quad A \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}. \tag{2.2}$$

◁

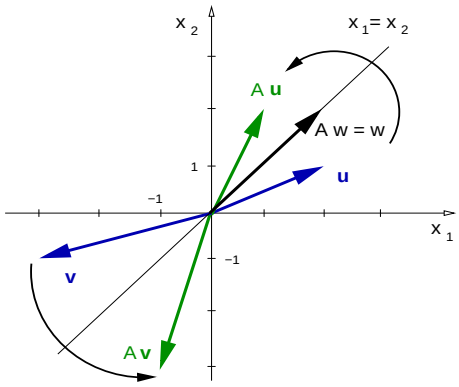

FIGURE 19. We sketch the action of matrix A in Example 2.1.3 on the vectors given in Eq. (2.2), which we called u, v, and w, respectively. Since $A : \mathbb{R}^2 \to \mathbb{R}^2$, we used the same plane to plot both u and Au.

EXAMPLE 2.1.4: Describe the function defined by the 2×2 matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

SOLUTION: The matrix A defines a function $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$, that can be interpreted as a *rotation by an angle $\pi/2$ counterclockwise*, see Fig. 20. Indeed, the action of the matrix A on an arbitrary element in $\mathsf{x} \in \mathbb{R}^2$ is the following,

$$\mathsf{Ax} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

Here are particular cases, represented in Fig. 20,

$$\mathsf{A} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \mathsf{A} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \end{bmatrix}, \quad \mathsf{A} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}. \tag{2.3}$$

◁

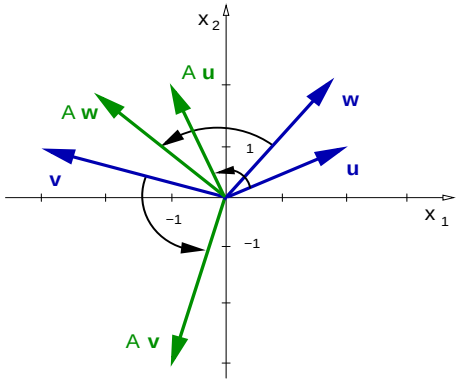

FIGURE 20. We sketch the action of matrix A in Example 2.1.3 on the vectors given in Eq. (2.2), which we called u, v, and w, respectively. Since $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$, we used the same plane to plot both u and Au.

EXAMPLE 2.1.5: Describe the function defined by the 2×2 matrix $\mathsf{A} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$, where θ is a fixed real number.

SOLUTION: The matrix A defines a function $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$, that can be interpreted as a *rotation by an angle θ counterclockwise*. To verify this, first compute $\mathsf{y} = \mathsf{Ax}$ for an arbitrary vector x, and then check that vector y is the counterclockwise rotation by θ of vector x.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \cos(\theta) - x_2 \sin(\theta) \\ x_1 \sin(\theta) + x_2 \cos(\theta) \end{bmatrix} \Rightarrow \begin{cases} y_1 = x_1 \cos(\theta) - x_2 \sin(\theta), \\ y_2 = x_1 \sin(\theta) + x_2 \cos(\theta). \end{cases}$$

We now show that the components y_1 and y_2 above are precisely the counterclockwise rotation by θ of the vector x. From Fig. 21 we see that the following relation holds:

$$y_1 = \cos(\theta + \phi)\, \|\mathsf{y}\|, \quad y_2 = \sin(\theta + \phi)\, \|\mathsf{y}\|,$$

where $\|\mathsf{y}\|$ is the magnitude of the vector y. Since a rotation does not change the magnitude of the vector, then $\|\mathsf{y}\| = \|\mathsf{x}\|$ and so,

$$y_1 = \cos(\theta + \phi)\, \|\mathsf{x}\|, \quad y_2 = \sin(\theta + \phi)\, \|\mathsf{x}\|.$$

Recalling now the formulas for the cosine and the sine of a sum of two angles,

$$\cos(\theta + \phi) = \cos(\theta)\, \cos(\phi) - \sin(\theta)\, \sin(\phi),$$
$$\sin(\theta + \phi) = \sin(\theta)\, \cos(\phi) + \cos(\theta)\, \sin(\phi),$$

we obtain that

$$y_1 = \cos(\theta)\,\cos(\phi)\,\|\mathsf{x}\| - \sin(\theta)\,\sin(\phi)\,\|\mathsf{x}\|,$$
$$y_2 = \sin(\theta)\,\cos(\phi)\,\|\mathsf{x}\| + \cos(\theta)\,\sin(\phi)\,\|\mathsf{x}\|.$$

Recalling that

$$x_1 = \cos(\phi)\,\|\mathsf{x}\|, \quad x_2 = \sin(\phi)\,\|\mathsf{x}\|,$$

we obtain the formula

$$y_1 = \cos(\theta)\,x_1 - \sin(\theta)\,x_2,$$
$$y_2 = \sin(\theta)\,x_1 + \sin(\theta)\,x_2.$$

This is precisely the expression that defines matrix A. Therefore, the action of the matrix A above is a rotation by θ counterclockwise. ◁

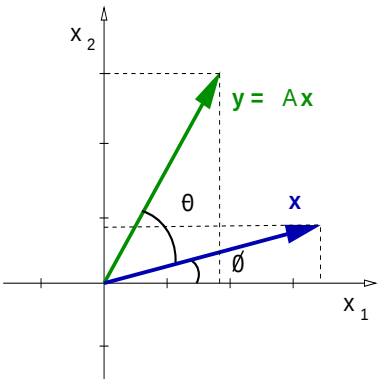

FIGURE 21. The vector y is the rotation by an angle θ counterclockwise of the vector x.

EXAMPLE 2.1.6: Describe the function defined by the 2×2 matrix $\mathsf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

SOLUTION: The matrix A defines a function $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$, which can be interpreted as a *dilation*, see Fig. 22. Indeed, the action of the matrix A on an arbitrary element in $\mathsf{x} \in \mathbb{R}^2$ is the following, $\mathsf{Ax} = 2\mathsf{x}$. ◁

EXAMPLE 2.1.7: Describe the function defined by the 2×2 matrix $\mathsf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

SOLUTION: The matrix A defines a function $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$, which can be interpreted as a *shear*, see Fig. 23. Indeed, the action of the matrix A on an arbitrary element in $\mathsf{x} \in \mathbb{R}^2$ is the following,

$$\mathsf{Ax} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_2 \end{bmatrix}.$$

◁

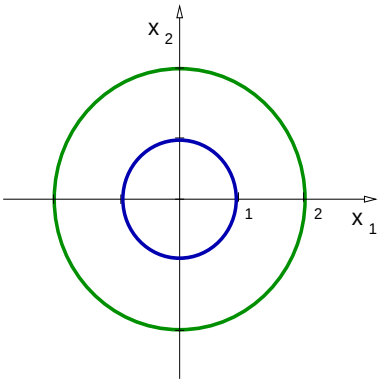

FIGURE 22. We sketch the action of matrix A in Example 2.1.6. Given any vector x with end point on the circle of radius one, the vector Ax = 2x is a vector parallel to x and with end point on the dashed circle of radius two.

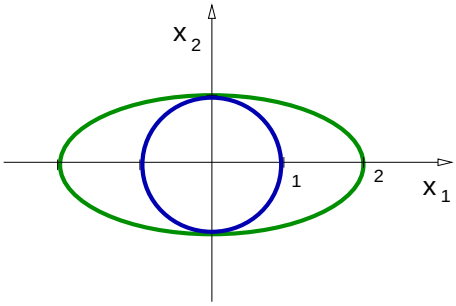

FIGURE 23. We sketch the action of matrix A in Example 2.1.7. Given any vector x with end point on the circle of radius one, the vector Ax = 2x is a vector parallel to x and with end point on the dashed curve.

2.1.2. **A matrix is a linear function.** We have seen several examples of functions given by matrices. All these functions are defined using the matrix-vector product. We have seen in Theorem 1.4.4 that the matrix-vector product is a linear operation. That is, given any $m \times n$ matrix A, for all vectors x, y $\in \mathbb{F}^n$ and all scalars a, $b \in \mathbb{F}$ holds that

$$A(a\mathsf{x} + b\mathsf{y}) = a\,A\mathsf{x} + b\,A\mathsf{y}$$

This property of the function defined by a matrix will be important later on, so any function with this property will be given a particular name.

Definition 2.1.2. *A function $T : \mathbb{F}^n \to \mathbb{F}^m$ is called a* ***linear transformation*** *iff for all vectors* x, y $\in \mathbb{F}^n$ *and for all scalars* a, $b \in \mathbb{F}$ *holds*

$$T(a\mathsf{x} + b\mathsf{y}) = a\,T(\mathsf{x}) + b\,T(\mathsf{y}).$$

The expression above contains the particular cases $a = b = 1$ and $b = 0$, which are respectively given by

$$T(\mathsf{x} + \mathsf{y}) = T(\mathsf{x}) + T(\mathsf{y}), \qquad T(a\mathsf{x}) = a\,T(\mathsf{x}).$$

We will also use the name *linear function* for a linear transformation. At the end of Section 2.2 we generalize the definition of a linear transformation to include functions on the

space of matrices. We will present two examples of linear functions on the space of matrices, the transpose and the trace functions. In this Section we present simpler examples of linear functions.

EXAMPLE 2.1.8: Show that the only linear transformations $T : \mathbb{R} \to \mathbb{R}$ are straight lines through the origin.

SOLUTION: Since $y = T(x)$ is linear and $x \in \mathbb{R}$, we have that

$$y = T(x) = T(x \times 1) = x\,T(1).$$

If we denote $T(1) = m$, we then conclude that a linear transformation $T : \mathbb{R} \to \mathbb{R}$ must be

$$y = mx,$$

which is a straight line through the origin with slope m. ◁

EXAMPLE 2.1.9: Any $m \times n$ matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ defines a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ by the equation $T(\mathsf{x}) = \mathsf{Ax}$. In particular, all the functions defined in Examples 2.1.1-2.1.7 are linear transformations. Consider the most general 2×2 matrix,

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathbb{F}^{2,2}$$

and explicitly show that the function $T(\mathsf{x}) = \mathsf{Ax}$ is a linear transformation.

SOLUTION: The explicit form of the function $T : \mathbb{F}^2 \to \mathbb{F}^2$ given by $T(\mathsf{x}) = \mathsf{Ax}$ is

$$T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \Rightarrow \quad T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 \\ A_{21}x_1 + A_{22}x_2 \end{bmatrix}.$$

This function is linear, as it can be seen from the following explicit computation,

$$\begin{aligned} T(c\mathsf{x} + d\mathsf{y}) &= T\Big(c \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + d \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\Big) \\ &= T\Big(\begin{bmatrix} cx_1 + dy_1 \\ cx_2 + dy_2 \end{bmatrix}\Big) \\ &= \begin{bmatrix} A_{11}(cx_1 + dy_1) + A_{12}(cx_2 + dy_2) \\ A_{21}(cx_1 + dy_1) + A_{22}(cx_2 + dy_2) \end{bmatrix} \\ &= \begin{bmatrix} A_{11}cx_1 + A_{12}cx_2 \\ A_{21}cx_1 + A_{22}cx_2 \end{bmatrix} + \begin{bmatrix} A_{11}dy_1 + A_{12}dy_2 \\ A_{21}dy_1 + A_{22}dy_2 \end{bmatrix} \\ &= c \begin{bmatrix} A_{11}x_1 + A_{12}x_2 \\ A_{21}x_1 + A_{22}x_2 \end{bmatrix} + d \begin{bmatrix} A_{11}y_1 + A_{12}y_2 \\ A_{21}y_1 + A_{22}y_2 \end{bmatrix} \\ &= c\,T(\mathsf{x}) + d\,T(\mathsf{y}). \end{aligned}$$

This establishes that T is a linear transformation. Notice that this proof is just a repetition of the Theorem 1.4.4 proof in the case of 2×2 matrices. ◁

EXAMPLE 2.1.10: Find a function $T : \mathbb{R}^2 \to \mathbb{R}^2$ that projects a vector onto the line $x_1 = x_2$, see Fig. 24. Show that this function is linear. Finally, find a matrix A such that $T(\mathsf{x}) = \mathsf{Ax}$.

SOLUTION: From Fig. 24 one can see that a possible way to compute the projection of a vector x onto the line $x_1 = x_2$ is the following: Add to the vector x its reflection along the line $x_1 = x_2$, and divide the result by two, that is,

$$T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \frac{1}{2}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}\Big) \quad \Rightarrow \quad T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \frac{(x_1 + x_2)}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We have obtained the projection function T. We now show that this function is linear: Indeed

$$\begin{aligned} T(a\mathsf{x}+b\mathsf{y}) &= T\Big(a\begin{bmatrix}x_1\\x_2\end{bmatrix}+b\begin{bmatrix}y_1\\y_2\end{bmatrix}\Big) \\ &= T\Big(\begin{bmatrix}ax_1+by_1\\ax_2+by_2\end{bmatrix}\Big) \\ &= \frac{(ax_1+by_1+ax_2+by_2)}{2}\begin{bmatrix}1\\1\end{bmatrix} \\ &= \frac{(ax_1+ax_2)}{2}\begin{bmatrix}1\\1\end{bmatrix}+\frac{(by_1+by_2)}{2}\begin{bmatrix}1\\1\end{bmatrix} \\ &= a\,T(\mathsf{x})+b\,T(\mathsf{y}). \end{aligned}$$

This shows that T is linear. We now find a matrix A such that $T(\mathsf{x})=\mathsf{A}\mathsf{x}$, as follows

$$\begin{aligned} T\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big) &= \frac{(x_1+x_2)}{2}\begin{bmatrix}1\\1\end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix}x_1+x_2\\x_1+x_2\end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \quad\Rightarrow\quad \mathsf{A}=\frac{1}{2}\begin{bmatrix}1&1\\1&1\end{bmatrix}. \end{aligned}$$

This matrix projects vectors onto the line $x_1=x_2$. ◁

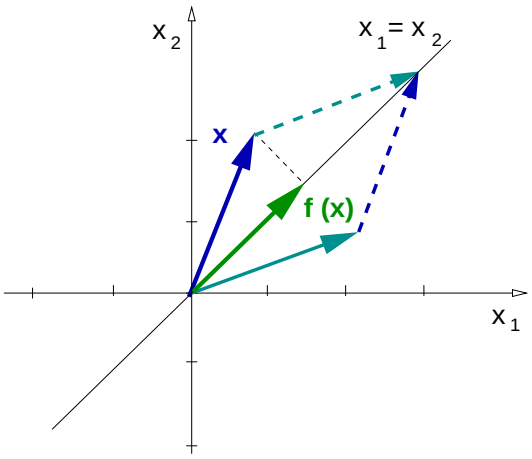

FIGURE 24. The function T projects the vector x onto the line $x_1=x_2$.

Further reading. See Sections 1.8 and 1.9 in Lay's book [2]. Also Sections 3.3 and 3.4 in Meyer's book [3].

2.1.3. **Exercises.**

2.1.1.- Determine which of the following functions $T : \mathbb{R}^2 \to \mathbb{R}^2$ is linear:

(a)
$$T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} 3x_2 \\ 2 + x_1 \end{bmatrix}.$$

(b)
$$T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}.$$

(c)
$$T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} x_1 x_2 \\ 0 \end{bmatrix}.$$

2.1.2.- Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $T(\mathsf{x}) = \mathsf{A}\mathsf{x}$, where
$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}.$$
Find $\mathsf{x} \in \mathbb{R}^2$ such that $T(\mathsf{x}) = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$.

2.1.3.- Given the matrix and vector
$$\mathsf{A} = \begin{bmatrix} 1 & -5 & -7 \\ -3 & 7 & 5 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} -2 \\ -2 \end{bmatrix},$$
define the function $T : \mathbb{R}^3 \to \mathbb{R}^2$ as $T(\mathsf{x}) = \mathsf{A}\mathsf{x}$, and then find all vectors x such that $T(\mathsf{x}) = \mathsf{b}$.

2.1.4.- Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that
$$T\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\Big) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad T\Big(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\Big) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$
Find the values of $T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big)$ for any vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbb{R}^2$.

2.1.5.- Describe geometrically what is the action of T over a vector $\mathsf{x} \in \mathbb{R}^2$, where

(a) $T(\mathsf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$;

(b) $T(\mathsf{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$;

(c) $T(\mathsf{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$;

(d) $T(\mathsf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

2.1.6.- Let $\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathsf{v} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$, $\mathsf{u} = \begin{bmatrix} 7 \\ -3 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation $T(\mathsf{x}) = x_1\mathsf{v} + x_2\mathsf{u}$. Find a matrix A such that $T(\mathsf{x}) = \mathsf{A}\mathsf{x}$ for all $\mathsf{x} \in \mathbb{R}^2$.

2.2. Linear combinations

One could say that the idea to introduce matrix operations originates from the interpretation of a matrix as a function. A matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ determines a function $\mathsf{A} : \mathbb{F}^n \to \mathbb{F}^m$. Such functions are generalizations of scalar valued functions of a single variable, $f : \mathbb{R} \to \mathbb{R}$. It is well-known how to compute the linear combination of two functions f, $g : \mathbb{R} \to \mathbb{R}$ and, when possible, how to compute their composition and their inverses. Matrices determine a particular generalizations of scalar functions where the operations mentioned above on scalar functions can be defined on matrices. The result is called the linear combination of matrices, and when possible, the product of matrices and the inverse of a matrix. Since matrices are generalizations of scalar valued functions, there are few operations on matrices that reduce to the identity operation in the the case of scalar functions. Among such operations belong the transpose of a matrix and the trace of a matrix.

2.2.1. **Linear combination of matrices.** The addition of two matrices and the multiplication of a matrix by scalar are defined component by component.

Definition 2.2.1. *Let* $\mathsf{A} = [A_{ij}]$ *and* $\mathsf{B} = [B_{ij}]$ *be* $m \times n$ *matrices in* $\mathbb{F}^{m,n}$ *and* a, $b \in \mathbb{F}$ *be any scalars. The* ***linear combination*** *of* A *and* B *is also and* $m \times n$ *matrix in* $\mathbb{F}^{m,n}$, *denoted as* $a\,\mathsf{A} + b\,\mathsf{B}$, *and given by*

$$a\,\mathsf{A} + b\,\mathsf{B} = [a\,A_{ij} + b\,B_{ij}].$$

Recall that the notation $a\mathsf{A} + b\mathsf{B} = [(aA + bB)_{ij}]$ means that the numbers $(aA + bB)_{ij}$ are the components of the matrix $a\mathsf{A} + b\mathsf{B}$. Using this notation the definition above can be expressed in terms of matrix components as follows

$$(aA + bB)_{ij} = a\,A_{ij} + b\,B_{ij}.$$

This definition contains the particular cases $a = b = 1$ and $b = 0$, given by, respectively,

$$(A + B)_{ij} = A_{ij} + B_{ij}, \qquad (aA)_{ij} = a\,A_{ij}.$$

Example 2.2.1: Consider the matrices

$$\mathsf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \qquad \mathsf{B} = \begin{bmatrix} 3 & 0 \\ 2 & -1 \end{bmatrix}.$$

(a) Find the matrix $\mathsf{A} + \mathsf{B}$ and $3\mathsf{A}$.
(b) Find a matrix C such that $2\mathsf{C} + 6\mathsf{A} = 4\mathsf{B}$.

Solution:
Part (a): The definition above gives,

$$\mathsf{A} + \mathsf{B} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 2 & -1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{A} + \mathsf{B} = \begin{bmatrix} 5 & -1 \\ 1 & 1 \end{bmatrix}, \qquad 3\mathsf{A} = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}.$$

Part (b): Matrix C is given by

$$\mathsf{C} = \frac{1}{2}\big(4\mathsf{B} - 6\mathsf{A}\big)$$

The definition above implies that

$$\mathsf{C} = 2\mathsf{B} - 3\mathsf{A} = 2\begin{bmatrix} 3 & 0 \\ 2 & -1 \end{bmatrix} - 3\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 4 & -2 \end{bmatrix} - 3\begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix},$$

therefore, we conclude that

$$\mathsf{C} = \begin{bmatrix} 0 & 3 \\ 7 & -8 \end{bmatrix}.$$

◁

We now summarize the main properties of the matrix linear combination.

Theorem 2.2.2. *For all matrices* $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{m,n}$ *and all scalars* $a, b \in \mathbb{F}$*, hold:*

(a) $(ab)\mathsf{A} = a(b\mathsf{A})$, *(associativity);*
(b) $a(\mathsf{A} + \mathsf{B}) = a\mathsf{A} + a\mathsf{B}$, *(distributivity);*
(c) $(a + b)\mathsf{A} = a\mathsf{A} + b\mathsf{A}$, *(distributivity);*
(d) $1\,\mathsf{A} = \mathsf{A}$, *(*$1 \in \mathbb{R}$ *is the identity).*

The definition of linear combination of matrices is defined as the linear combination of their components, which are real of complex numbers. Therefore, all properties in Theorem 2.2.2 on linear combinations of matrices are obtained from the analogous properties on the linear combination of real or complex numbers.

Proof of Theorem 2.2.2: We use components notation.

(a):
$$\big[(ab)A\big]_{ij} = (ab)A_{ij} = a(bA_{ij}) = a(bA)_{ij} = \big[a(bA)\big]_{ij}.$$

(b):
$$\big[a(A+B)\big]_{ij} = a(A+B)_{ij} = a(A_{ij} + B_{ij}) = aA_{ij} + aB_{ij} = (aA)_{ij} + (aB)_{ij} = \big[aA + aB\big]_{ij}.$$

(c):
$$\big[(a+b)A\big]_{ij} = (a+b)A_{ij} = aA_{ij} + bA_{ij} = (aA)_{ij} + (bA)_{ij} = \big[aA + bA\big]_{ij}.$$

(d):
$$(1A)_{ij} = 1\,A_{ij} = A_{ij}.$$

□

2.2.2. **The transpose, adjoint, and trace of a matrix.** Since matrices are generalizations of scalar-valued functions, one can define operations on matrices that, unlike linear combinations, have no analogue on scalar-valued functions. One of such operations is the transpose of a matrix, which is a new matrix with the rows and columns interchanged.

Definition 2.2.3. *The* ***transpose*** *of a matrix* $\mathsf{A} = [A_{ij}] \in \mathbb{F}^{m,n}$ *is the matrix denoted as* $\mathsf{A}^T = \big[(A^T)_{kl}\big] \in \mathbb{F}^{n,m}$*, with its components given by*

$$\big(A^T\big)_{kl} = A_{lk}.$$

EXAMPLE 2.2.2: Find the transpose of the 2×3 matrix $\mathsf{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$.

SOLUTION: Matrix A has components A_{ij} with $i = 1, 2$ and $j = 1, 2, 3$. Therefore, its transpose has components $(A^T)_{ji} = A_{ij}$, that is, A^T has three rows and two columns,

$$\mathsf{A}^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

◁

EXAMPLE 2.2.3: Show that the transpose operation satisfies $(\mathsf{A}^T)^T = \mathsf{A}$.

SOLUTION: The proof is:
$$\big[(A^T)^T\big]_{ij} = (A^T)_{ji} = A_{ij}.$$
An example of this property is the following: In Example 2.2.2 we showed that

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

Therefore,

$$\left(\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}^T \right)^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}.$$

◁

If a matrix has complex-valued coefficients, then the complex conjugate of a matrix, also called the conjugate, is defined as the conjugate of each component.

Definition 2.2.4. *The* ***complex conjugate*** *of a matrix* $\mathsf{A} = [A_{ij}] \in \mathbb{F}^{m,n}$ *is the matrix*

$$\overline{\mathsf{A}} = \left[\overline{A}_{ij} \right] \in \mathbb{F}^{m,n}.$$

EXAMPLE 2.2.4: Find the conjugate if matrix $\mathsf{A} = \begin{bmatrix} 1 & 2+i \\ -i & 3-4i \end{bmatrix}$.

SOLUTION: We have to conjugate each component of matrix A, that is,

$$\overline{\mathsf{A}} = \begin{bmatrix} 1 & 2-i \\ i & 3+4i \end{bmatrix}.$$

◁

EXAMPLE 2.2.5: Show that a matrix A has real coefficients iff $\mathsf{A} = \overline{\mathsf{A}}$; and a matrix has purely imaginary coefficients iff $\mathsf{A} = -\overline{\mathsf{A}}$. Show one example of each case.

SOLUTION: The first condition is $\mathsf{A} = \overline{\mathsf{A}}$, that is, $A_{ij} = \overline{A}_{ij}$ holds for every matrix component, which implies $2i\,\mathrm{Im}(A_{ij}) = A_{ij} - \overline{A}_{ij} = 0$. Therefore, every matrix component A_{ij} is real. The second condition is $\mathsf{A} = -\overline{\mathsf{A}}$, that is, $A_{ij} = -\overline{A}_{ij}$ holds for every matrix component, which implies $2\mathrm{Re}(A_{ij}) = A_{ij} + \overline{A}_{ij} = 0$. Therefore, every matrix component A_{ij} is purely imaginary. Here are examples of these two situations:

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \Rightarrow \quad \overline{\mathsf{A}} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \mathsf{A};$$

$$\mathsf{A} = \begin{bmatrix} i & 2i \\ 3i & 4i \end{bmatrix} \quad \Rightarrow \quad \overline{\mathsf{A}} = \begin{bmatrix} -i & -2i \\ -3i & -4i \end{bmatrix} = -\mathsf{A}.$$

◁

Definition 2.2.5. *The* ***adjoint*** *of a matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *is the matrix* $\mathsf{A}^* = \overline{\mathsf{A}}^T \in \mathbb{F}^{n,m}$.

It is not difficult to show, using components, that $\left(\overline{\mathsf{A}}\right)^T = \overline{(\mathsf{A}^T)}$, that is, the order of the transpose an conjugate operation does not change the resulting matrix. This property is the reason why there is no parenthesis in the definition of A^*.

EXAMPLE 2.2.6: Find the adjoint of matrix $\mathsf{A} = \begin{bmatrix} 1 & 2+i \\ -i & 3-4i \end{bmatrix}$.

SOLUTION: We need to switch rows with columns and complex conjugate the result, that is,

$$\mathsf{A}^* = \begin{bmatrix} 1 & i \\ 2-i & 3+4i \end{bmatrix}.$$

◁

The transpose, conjugate and adjoint operations are useful to specify different matrix classes having particular symmetries. These matrix classes are the symmetric, the skew-symmetric, the Hermitian, and the skew-Hermitian matrices. Here is a precise definition.

Definition 2.2.6. *An $n \times n$ matrix A is called:*

(a) ***symmetric*** *iff holds $\mathsf{A} = \mathsf{A}^T$;*
(b) ***skew-symmetric*** *iff holds $\mathsf{A} = -\mathsf{A}^T$;*
(c) ***Hermitian*** *iff holds $\mathsf{A} = \mathsf{A}^*$;*
(d) ***skew-Hermitian*** *iff holds $\mathsf{A} = -\mathsf{A}^*$.*

EXAMPLE 2.2.7: We present examples of each of the classes introduced in Def. 2.2.6.

Part (a): Matrices A and B are symmetric. Notice that A is also Hermitian, while B is not Hermitian,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 7 & 4 \\ 3 & 4 & 8 \end{bmatrix} = \mathsf{A}^T, \qquad \mathsf{B} = \begin{bmatrix} 1 & 2+3i & 3 \\ 2+3i & 7 & 4i \\ 3 & 4i & 8 \end{bmatrix} = \mathsf{B}^T.$$

Part (b): Matrix C is skew-symmetric,

$$\mathsf{C} = \begin{bmatrix} 0 & -2 & 3 \\ 2 & 0 & -4 \\ -3 & 4 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{C}^T = \begin{bmatrix} 0 & 2 & -3 \\ -2 & 0 & 4 \\ 3 & -4 & 0 \end{bmatrix} = -\mathsf{C}.$$

Notice that the diagonal elements in a skew-symmetric matrix must vanish, since $C_{ij} = -C_{ji}$ in the case $i = j$ means $C_{ii} = -C_{ii}$, that is, $C_{ii} = 0$.

Part (c): Matrix D is Hermitian but is not symmetric:

$$\mathsf{D} = \begin{bmatrix} 1 & 2+i & 3 \\ 2-i & 7 & 4+i \\ 3 & 4-i & 8 \end{bmatrix} \quad \Rightarrow \quad \mathsf{D}^T = \begin{bmatrix} 1 & 2-i & 3 \\ 2+i & 7 & 4-i \\ 3 & 4+i & 8 \end{bmatrix} \neq \mathsf{D},$$

however,

$$\mathsf{D}^* = \overline{\mathsf{D}}^T = \begin{bmatrix} 1 & 2+i & 3 \\ 2-i & 7 & 4+i \\ 3 & 4-i & 8 \end{bmatrix} = \mathsf{D}.$$

Notice that the diagonal elements in a Hermitian matrix must be real numbers, since the condition $A_{ij} = \overline{A}_{ji}$ in the case $i = j$ implies $A_{ii} = \overline{A}_{ii}$, that is, $2i\text{Im}(A_{ii}) = A_{ii} - \overline{A}_{ii} = 0$. We can also verify what we said in part (a), matrix A is Hermitian since $\mathsf{A}^* = \overline{\mathsf{A}}^T = \mathsf{A}^T = \mathsf{A}$.

Part (d): The following matrix E is skew-Hermitian:

$$\mathsf{E} = \begin{bmatrix} i & 2+i & -3 \\ -2+i & 7i & 4+i \\ 3 & -4+i & 8i \end{bmatrix} \quad \Rightarrow \quad \mathsf{E}^T = \begin{bmatrix} i & -2+i & 3 \\ 2+i & 7i & -4+i \\ -3 & 4+i & 8i \end{bmatrix}$$

therefore,

$$\mathsf{E}^* = \overline{\mathsf{E}}^T \begin{bmatrix} -i & -2-i & 3 \\ 2-i & -7i & -4-i \\ -3 & 4-i & -8i \end{bmatrix} = -\mathsf{E}.$$

A skew-Hermitian matrix has purely imaginary elements in its diagonal, and the off diagonal elements have skew-symmetric real parts with symmetric imaginary parts. ◁

The trace of a square matrix is a number, the sum of all the diagonal matrix coefficients.

Definition 2.2.7. *The* ***trace*** *of a square matrix $\mathsf{A} = [A_{ij}] \in \mathbb{F}^{n,n}$, denoted as $\text{tr}(\mathsf{A}) \in \mathbb{F}$, is the scalar given by*

$$\text{tr}(\mathsf{A}) = A_{11} + \cdots + A_{nn}.$$

EXAMPLE 2.2.8: Find the trace of the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

SOLUTION: We add up the diagonal elements: $\text{tr}\,(\mathsf{A}) = 1 + 5 + 9$, that is, $\text{tr}\,(\mathsf{A}) = 15$. ◁

2.2.3. **Linear transformations on matrices.** The operation of computing the transpose matrix can be thought as function on the set of all $m \times n$ matrices. The transpose operation is in fact a function $T : \mathbb{F}^{m,n} \to \mathbb{F}^{n,m}$ given by $T(\mathsf{A}) = \mathsf{A}^T$. In a similar way, the operation of computing the trace of a square matrix can be thought as a function on the set of all square matrices. The trace is a function $\text{tr} : \mathbb{F}^{n,n} \to \mathbb{F}$, where $\text{tr}\,(\mathsf{A}) = A_{11} + \cdots + A_{nn}$. One can verify that transpose function T and the trace function tr are indeed linear functions.

Theorem 2.2.8. *Both the transpose function $T : \mathbb{F}^{m,n} \to \mathbb{F}^{n,m}$, given by $\mathbb{T}(\mathsf{A}) = \mathsf{A}^T$, and the trace function $\text{tr} : \mathbb{F}^{n,n} \to \mathbb{F}$, given by $\text{tr}\,(\mathsf{A}) = A_{11} + \cdots + A_{nn}$, are linear functions.*

Proof of Theorem 2.2.8: We start with the transpose function $T(\mathsf{A}) = \mathsf{A}^T$, which in matrix components has the form $\big(T(\mathsf{A})\big)_{ij} = A_{ji}$. Therefore, given two matrices $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{m,n}$ and arbitrary scalars $a, b \in \mathbb{F}$, holds.

$$\begin{aligned} \big(T(a\mathsf{A} + b\mathsf{B})\big)_{ij} &= \big(a\mathsf{A} + b\mathsf{B}\big)_{ji} \\ &= a\,\mathsf{A}_{ji} + b\,\mathsf{B}_{ji} \\ &= a\,\big(T(\mathsf{A})\big)_{ij} + b\,\big(T(\mathsf{B})\big)_{ij}, \end{aligned}$$

so we conclude that

$$T(a\mathsf{A} + b\mathsf{B}) = a\,T(\mathsf{A}) + b\,T(\mathsf{B}),$$

showing that the transpose operation is a linear function. We now consider the trace function. Given any two matrices all $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{n,n}$ and arbitrary scalars $a, b \in \mathbb{F}$, holds,

$$\begin{aligned} \text{tr}\,(a\mathsf{A} + b\mathsf{B}) &= \sum_{i=1}^{n} \big(a\mathsf{A} + b\mathsf{B}\big)_{ii} \\ &= \sum_{i=1}^{n} \big(a\,\mathsf{A}_{ii} + b\,\mathsf{B}_{ii}\big) \\ &= a \sum_{i=1}^{n} \mathsf{A}_{ii} + b \sum_{i=1}^{n} \mathsf{B}_{ii} \\ &= a\,\text{tr}\,(\mathsf{A}) + b\,\text{tr}\,(\mathsf{B}). \end{aligned}$$

This shows that the trace is a linear function. This establishes the Theorem. □

2.2.4. **Exercises.**

2.2.1.- Construct an example of a 3×3 matrix satisfying:

(a) Is symmetric and skew-symmetric.

(b) Is Hermitian and symmetric.

(c) Is Hermitian but not symmetric.

2.2.2.- Find the numbers x, y, z solution of the equation

$$2 \begin{bmatrix} x+2 & y+3 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ y & z \end{bmatrix}^T$$

2.2.3.- Given any square matrix A show that $\mathsf{A}+\mathsf{A}^T$ is a symmetric matrix, while $\mathsf{A}-\mathsf{A}^T$ is a skew-symmetric matrix.

2.2.4.- Prove that there is only one way to express a matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ as a sum of a symmetric matrix and a skew-symmetric matrix.

2.2.5.- Prove the following statements:

(a) If $\mathsf{A} = [A_{ij}]$ is a skew-symmetric matrix, then holds $A_{ii} = 0$.

(b) If $\mathsf{A} = [A_{ij}]$ is a skew-Hermitian matrix, then the non-zero coefficients A_{ii} are purely imaginary.

(c) If A is a real and symmetric matrix, then $\mathsf{B} = i\mathsf{A}$ is skew-Hermitian.

2.2.6.- Prove that for all A, $\mathsf{B} \in \mathbb{F}^{m,n}$ and all a, $b \in \mathbb{F}$ holds

$$(a\mathsf{A} + b\mathsf{B})^* = \overline{a}\,\mathsf{A}^* + \overline{b}\,\mathsf{B}^*.$$

2.2.7.- Prove that the transpose function $T : \mathbb{F}^{m,n} \to \mathbb{F}^{n,m}$ and trace function $\text{tr} : \mathbb{F}^{n,n} \to \mathbb{F}$ are linear functions.

2.3. Matrix multiplication

The operation of matrix multiplication originates in the composition of functions. We call this operation a multiplication since it reduces to the multiplication of real numbers in the case of 1×1 real matrices. Unlike the multiplication of real numbers, the product of general matrices is not commutative, that is, $\mathsf{AB} \neq \mathsf{BA}$ in the general case. This property reflects the fact that the composition of two functions is a non-commutative operation. In this Subsection we first introduce the multiplication of two matrices using the matrix vector product. We then introduce the formula for the components of the product of two matrices. Finally we show that the composition of two matrices is their matrix product.

2.3.1. **Algebraic definition.** Matrix multiplication is defined using the matrix-vector product. The product of two matrices is the matrix-vector product of one matrix with each column of the other matrix.

Definition 2.3.1. *The* ***matrix multiplication*** *of the $m \times n$ matrix A with the $n \times \ell$ matrix $\mathsf{B} = \big[\mathsf{B}_{:1}, \cdots, \mathsf{B}_{:\ell}\big]$, denoted by AB, is the $m \times \ell$ matrix given by $\mathsf{AB} = \big[\mathsf{AB}_{:1}, \cdots, \mathsf{AB}_{:\ell}\big]$.*

The product is not defined for two arbitrary matrices, since the size of the matrices is important: The numbers of columns in the first matrix must match the numbers of rows in the second matrix,

$$\begin{array}{ccccc} \mathsf{A} & \text{times} & \mathsf{B} & \text{defines} & \mathsf{AB} \\ m \times \boldsymbol{n} & & \boldsymbol{n} \times \ell & & m \times \ell \end{array}$$

We assign a name to matrices satisfying this property.

Definition 2.3.2. *Matrices A and B are called* ***conformable*** *in the order A, B, iff the product AB is well defined.*

Example 2.3.1: Compute the product of the matrices A and B below, which are conformable in both orders AB and BA, where

$$\mathsf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} 3 & 0 \\ 2 & -1 \end{bmatrix}.$$

Solution: Following the definition above we compute the product in the order AB, namely,

$$\mathsf{AB} = \big[\mathsf{AB}_{:1}, \mathsf{AB}_{:2}\big] = \left[\mathsf{A}\begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathsf{A}\begin{bmatrix} 0 \\ -1 \end{bmatrix}\right] = \left[\begin{bmatrix} 6-2 \\ -3+4 \end{bmatrix}, \begin{bmatrix} 0+1 \\ 0-2 \end{bmatrix}\right] \Rightarrow \mathsf{AB} = \begin{bmatrix} 4 & 1 \\ 1 & -2 \end{bmatrix}.$$

Using the same definition we can compute the product in the opposite order, that is,

$$\mathsf{BA} = \big[\mathsf{BA}_{:1}, \mathsf{BA}_{:2}\big] = \left[\mathsf{B}\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathsf{B}\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right] = \left[\begin{bmatrix} 6-0 \\ 4+1 \end{bmatrix}, \begin{bmatrix} -3+0 \\ -2-2 \end{bmatrix}\right] \Rightarrow \mathsf{BA} = \begin{bmatrix} 6 & -3 \\ 5 & -4 \end{bmatrix}.$$

This is an example where we have that $\mathsf{AB} \neq \mathsf{BA}$. ◁

The following result gives a formula to compute the components of the product matrix in terms of the components of the individual matrices.

Theorem 2.3.3. *Consider the $m \times n$ matrix $\mathsf{A} = [A_{ij}]$ and the $n \times \ell$ matrix $\mathsf{B} = [B_{jk}]$, where the indices take values as follows: $i = 1, \cdots, m$, $j = 1, \cdots, n$ and $k = 1, \cdots, \ell$. The components of the product matrix AB are given by*

$$(AB)_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk}. \tag{2.4}$$

We recall that the symbol $\sum_{j=1}^{n}$ in Eq. (2.4) means to add up all the terms having the index j starting from $j = 1$ until $j = n$, that is,

$$\sum_{j=1}^{n} A_{ij}B_{jk} = A_{i1}B_{1k} + A_{i2}B_{2k} + \cdots + A_{in}B_{nk}.$$

Proof of Theorem 2.3.3: The column k of the product AB is given by the column vector $\mathsf{AB}_{:k}$. This is a vector with m components,

$$\mathsf{AB}_{:k} = \begin{bmatrix} \sum_{j=1}^{n} A_{1j}B_{jk} \\ \vdots \\ \sum_{j=1}^{n} A_{mj}B_{jk} \end{bmatrix},$$

therefore the i-th component of this vector is given by

$$(AB)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk}.$$

This establishes the Theorem. □

EXAMPLE 2.3.2: We now use Eq. (2.4) to find the product of matrices A and B in Example 2.3.1. The component $(AB)_{11} = 4$ is obtained from the first row in matrix A and the first column in matrix B as follows:

$$\begin{bmatrix} \mathbf{2} & \mathbf{-1} \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{3} & 0 \\ \mathbf{2} & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{4} & 1 \\ 1 & -2 \end{bmatrix}, \qquad (2)(3) + (-1)(2) = 4;$$

The component $(AB)_{12} = -1$ is obtained as follows:

$$\begin{bmatrix} \mathbf{2} & \mathbf{-1} \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 & \mathbf{0} \\ 2 & \mathbf{-1} \end{bmatrix} = \begin{bmatrix} 4 & \mathbf{1} \\ 1 & -2 \end{bmatrix}, \qquad (2)(0) + (-1)(1) = -1;$$

The component $(AB)_{21} = 1$ is obtained as follows:

$$\begin{bmatrix} 2 & -1 \\ \mathbf{-1} & \mathbf{2} \end{bmatrix} \begin{bmatrix} \mathbf{3} & 0 \\ \mathbf{2} & -1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ \mathbf{1} & -2 \end{bmatrix}, \qquad (-1)(3) + (2)(2) = 1;$$

And finally the component $(AB)_{22} = -2$ is obtained as follows:

$$\begin{bmatrix} 2 & -1 \\ \mathbf{-1} & \mathbf{2} \end{bmatrix} \begin{bmatrix} 3 & \mathbf{0} \\ 2 & \mathbf{-1} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 1 & \mathbf{-2} \end{bmatrix}, \qquad (-1)(0) + (2)(-1) = -2.$$

◁

We have seen in Example 2.3.1 that the matrix product is not commutative, since in that example $\mathsf{AB} \neq \mathsf{BA}$. In that example the matrices were conformable in both orders AB and BA, although their products do not match. It can also be possible that two matrices A and B are conformable in the order AB but they are not conformable in the opposite order. That is, the matrix product is possible in one order but not in the other order.

EXAMPLE 2.3.3: Consider the matrices

$$\mathsf{A} = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \qquad \mathsf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

These matrices are conformable in the order AB but not in the order BA. In the first case we obtain

$$\mathsf{AB} = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \Rightarrow \quad \mathsf{AB} = \begin{bmatrix} 16 & 23 & 30 \\ 6 & 9 & 12 \end{bmatrix}.$$

The product BA is not possible. ◁

EXAMPLE 2.3.4: Column vectors and row vectors are particular cases of matrices, they are $n \times 1$ and $1 \times n$ matrices, respectively. We denote row vectors as transpose of a column vector. Using this notation and the matrix product, compute both products $v^T u$ and $u\, v^T$, where the vectors are given by

$$u = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \qquad v = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

SOLUTION: In the first case we multiply the matrices 1×2 and 2×1, so the result is a 1×1 matrix, a real number, given by,

$$v^T u = \begin{bmatrix} 5 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \Rightarrow \quad v^T u = 13.$$

In the second case we multiply the matrices 2×1 and 1×2, so the result is a 2×2 matrix,

$$u\, v^T = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \begin{bmatrix} 5 & 1 \end{bmatrix} \quad \Rightarrow \quad u\, v^T = \begin{bmatrix} 10 & 2 \\ 15 & 3 \end{bmatrix}.$$

◁

It is well-known that the product of two numbers is zero, then one of them must be zero. This property is not true in the case of matrix multiplication, as it can be seen below.

EXAMPLE 2.3.5: Compute the product AB where $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$.

SOLUTION: It is simple to check that

$$AB = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad AB = 0.$$

The product is the zero matrix, although $A \neq 0$ and $B \neq 0$. ◁

2.3.2. **Matrix composition.** The product of two matrices originates in the composition of the linear functions defined by the matrices. This can seen in the following result.

Theorem 2.3.4. *Given the $m \times n$ matrix $A : \mathbb{F}^n \to \mathbb{F}^m$ and the $n \times \ell$ matrix $B : \mathbb{F}^\ell \to \mathbb{F}^n$, their composition is a function*

$$A \circ B : \mathbb{F}^\ell \xrightarrow{B} \mathbb{F}^n \xrightarrow{A} \mathbb{F}^m,$$

which is an $m \times \ell$ matrix given by the matrix product of A and B, that is, $A \circ B = AB$.

Proof of Theorem 2.3.4: The composition of the function A and B is defined for all $x \in \mathbb{F}^\ell$ as follows

$$(A \circ B)\boldsymbol{x} = A(Bx).$$

Introduce the usual notation $B = [B_{:1}, \cdots, B_{:\ell}]$. Then, the composition $A \circ B$ can be re-expressed as follows,

$$(A \circ B)\boldsymbol{x} = A\Big(\big[B_{:1}, \cdots, B_{:\ell}\big] \begin{bmatrix} x_1 \\ \vdots \\ x_\ell \end{bmatrix} \Big) = A\big(B_{:1}x_1 + \cdots + B_{:\ell}x_\ell\big).$$

Since the matrix-vector product is a linear operation, we get

$$(A \circ B)\boldsymbol{x} = A\big(B_{:1}x_1\big) + \cdots + A\big(B_{:\ell}x_\ell\big) = \big(AB_{:1}\big)x_1 + \cdots + \big(AB_{:\ell}\big)x_\ell$$

where the second equation comes from the definition of matrix multiplication. Then, using again the definition of matrix-vector product,

$$(\mathsf{A}\circ\mathsf{B})\boldsymbol{x} = \big[\mathsf{AB}_{:1},\cdots,\mathsf{AB}_{:\ell}\big]\begin{bmatrix} x_1 \\ \vdots \\ x_\ell \end{bmatrix} \quad \Rightarrow \quad (\mathsf{A}\circ\mathsf{B})\boldsymbol{x} = (\mathsf{AB})\mathsf{x}.$$

This establishes the Theorem. □

EXAMPLE 2.3.6: Find the matrix $\mathsf{T}:\mathbb{R}^2\to\mathbb{R}^2$ that produces a rotation by an angle θ_1 counterclockwise and then another rotation by an angle θ_2 counterclockwise.

SOLUTION: Let us denote by R_θ the matrix that performs a rotation on the plane by an angle θ counterclockwise, that is,

$$\mathsf{R}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The matrix T is the the composition $\mathsf{T} = \mathsf{R}_{\theta_2}\circ\mathsf{R}_{\theta_1}$. Since the matrix of a composition of these two rotations can be obtained computing the matrix product of them, then we obtain

$$\mathsf{T} = \mathsf{R}_{\theta_2}\mathsf{R}_{\theta_1} = \begin{bmatrix} \cos(\theta_2) & -\sin(\theta_2) \\ \sin(\theta_2) & \cos(\theta_2) \end{bmatrix}\begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) \\ \sin(\theta_1) & \cos(\theta_1) \end{bmatrix}.$$

The product above is given by

$$\mathsf{T} = \begin{bmatrix} \cos(\theta_2)\cos(\theta_1) - \sin(\theta_2)\sin(\theta_1) & -\big[\sin(\theta_1)\cos(\theta_2) + \sin(\theta_2)\cos(\theta_1)\big] \\ \sin(\theta_1)\cos(\theta_2) + \sin(\theta_2)\cos(\theta_1) & -\sin(\theta_2)\sin(\theta_1) + \cos(\theta_2)\cos(\theta_1) \end{bmatrix}.$$

The formulas

$$\cos(\theta_1+\theta_2) = \cos(\theta_2)\cos(\theta_1) - \sin(\theta_2)\sin(\theta_1)$$
$$\sin(\theta_1+\theta_2) = \sin(\theta_1)\cos(\theta_2) + \sin(\theta_2)\cos(\theta_1),$$

imply that

$$\mathsf{T} = \begin{bmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{bmatrix}.$$

Notice that we have obtained a result that it is intuitively clear: Two consecutive rotations on the plane is equivalent to a single rotation by an angle that is the sum of the individual rotations, that is,

$$\mathsf{R}_{\theta_2}\mathsf{R}_{\theta_1} = \mathsf{R}_{\theta_2+\theta_1}.$$

In particular, notice that the matrix multiplication when restricted to the set of all rotation matrices on the plane is a commutative operation, that is,

$$\mathsf{R}_{\theta_2}\mathsf{R}_{\theta_1} = \mathsf{R}_{\theta_1}\mathsf{R}_{\theta_2}, \qquad \theta_1,\theta_2\in\mathbb{R}.$$

◁

In Examples 2.3.7 and 2.3.8 below we show that the order of the functions in a composition change the resulting function. This is essentially the reason behind the non-commutativity of the matrix multiplication.

EXAMPLE 2.3.7: Find the matrix $\mathsf{T}:\mathbb{R}^2\to\mathbb{R}^2$ that first performs a rotation by an angle $\pi/2$ counterclockwise and then performs a reflection along the $x_1 = x_2$ line on the plane.

SOLUTION: The matrix T is the composition the rotation $\mathsf{R}_{\pi/2}$ with the reflection function A, given by

$$\mathsf{R}_{\pi/2} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad \mathsf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The matrix T is given by

$$\mathsf{T} = \mathsf{A}\mathsf{R}_{\pi/2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{T} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Therefore, the function T is a reflection along the horizontal line $x_2 = 0$. ◁

EXAMPLE 2.3.8: Find the matrix $\mathsf{S} : \mathbb{R}^2 \to R^2$ that first performs a reflection along the $x_1 = x_2$ line on the plane an then performs a rotation by an angle $\pi/2$ counterclockwise.

SOLUTION: The matrix S is the composition the reflection function A and then the rotation $\mathsf{R}_{\pi/2}$, given in the previous Example 2.3.7. The matrix S is the given by

$$\mathsf{S} = \mathsf{R}_{\pi/2}\mathsf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{S} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Therefore, the function S is a reflection along the vertical line $x_1 = 0$. ◁

2.3.3. **Main properties.** We summarize the main properties of the matrix multiplication.

Theorem 2.3.5. *The following properties hold for all $m \times n$ matrix A, $n \times m$ matrix B, $n \times \ell$ matrices C, D, and $\ell \times k$ matrix E:*

(a) $\mathsf{AB} \neq \mathsf{BA}$ *in the general case, so the product is non-commutative;*
(b) $\mathsf{A}(\mathsf{C} + \mathsf{D}) = \mathsf{AC} + \mathsf{AD}$, *and* $(\mathsf{C} + \mathsf{D})\mathsf{E} = \mathsf{CE} + \mathsf{DE}$;
(c) $\mathsf{A}(\mathsf{CE}) = (\mathsf{AC})\mathsf{E}$, *associativity;*
(d) $\mathsf{I}_m\mathsf{A} = \mathsf{A}\mathsf{I}_n = \mathsf{A}$;
(e) $(\mathsf{AC})^T = \mathsf{C}^T\mathsf{A}^T$, *and* $(\mathsf{AC})^* = \mathsf{C}^*\mathsf{A}^*$.
(f) $\operatorname{tr}(\mathsf{AB}) = \operatorname{tr}(\mathsf{BA})$.

Proof of Theorem 2.3.5:
Part (a): When $m \neq n$ the matrix AB is $m \times m$ while BA is $n \times n$, so they cannot be equal. When $m = n$, the matrices in Example 2.3.4 and 2.3.5 show that this product is not commutative.
Part (b): This property can be shown as follows:

$$\begin{aligned} \mathsf{A}(\mathsf{C} + \mathsf{D}) &= \mathsf{A}\big(\big[\mathsf{C}_{:1}, \cdots, \mathsf{C}_{:\ell}\big] + \big[\mathsf{D}_{:1}, \cdots, \mathsf{D}_{:\ell}\big]\big) \\ &= \mathsf{A}\left[(\mathsf{C}_{:1} + \mathsf{D}_{:1}), \cdots, (\mathsf{C}_{:\ell} + \mathsf{D}_{:\ell})\right] \\ &= \left[\mathsf{A}(\mathsf{C}_{:1} + \mathsf{D}_{:1}), \cdots, \mathsf{A}(\mathsf{C}_{:\ell} + \mathsf{D}_{:\ell})\right] \\ &= \left[(\mathsf{AC}_{:1} + \mathsf{AD}_{:1}), \cdots, (\mathsf{AC}_{:\ell} + \mathsf{AD}_{:\ell})\right] \\ &= \left[\mathsf{AC}_{:1}, \cdots, \mathsf{AC}_{:\ell}\right] + \left[\mathsf{AD}_{:1}, \cdots, \mathsf{AD}_{:\ell}\right] \\ &= \mathsf{AC} + \mathsf{AD}. \end{aligned}$$

The other equation is proven in a similar way.
Part (c): This property is proven using the component expression for the matrix product:

$$\big[A(CE)\big]_{ij} = \sum_{k=1}^{n} A_{ik}(CE)_{kj} = \sum_{k=1}^{n} A_{ik}\Big(\sum_{l=1}^{\ell} C_{kl}E_{lj}\Big);$$

however, the order of the sums can be interchanged,

$$\sum_{k=1}^{n} A_{ik}\Big(\sum_{l=1}^{\ell} C_{kl}E_{lj}\Big) = \sum_{l=1}^{\ell}\Big(\sum_{k=1}^{n} A_{ik}C_{kl}\Big)E_{lj};$$

So, from this last expression is not difficult to show that:

$$\sum_{l=1}^{\ell}\Big(\sum_{k=1}^{n} A_{ik}C_{kl}\Big)E_{lj} = \sum_{l=1}^{\ell}(AC)_{il}E_{lj} = \big[(AC)E\big]_{ij};$$

We have just proved that

$$\big[A(CE)\big]_{ij} = \big[(AC)E\big]_{ij}.$$

Part (d): We can use components again, recalling that the components of I_m are given by $(I_m)_{ij} = 0$ if $i \neq j$ and is $(I_m)_{ii} = 1$. Therefore,

$$(I_m A)_{ij} = \sum_{k=1}^{m}(I_m)_{ik}A_{kj} = (I_m)_{ii}A_{ij} = A_{ij}.$$

Analogously,

$$(AI_n)_{ij} = \sum_{k=1}^{n} A_{ik}(I_n)_{kj} = A_{ij}(I_n)_{jj} = A_{ij}.$$

Part (e): Use components once more:

$$\big[(AC)^T\big]_{ij} = (AC)_{ji} = \sum_{k=1}^{n} A_{jk}C_{ki} = \sum_{k=1}^{n}(A^T)_{kj}(C^T)_{ik} = \sum_{k=1}^{n}(C^T)_{ik}(A^T)_{kj} = \big[C^T A^T\big]_{ij}.$$

The second equation follows from the proof above and the property of the complex conjugate:

$$\overline{(\mathsf{AC})} = \overline{\mathsf{A}}\,\overline{\mathsf{C}}.$$

Indeed

$$(\mathsf{AC})^* = \overline{(\mathsf{AC})^T} = \overline{\mathsf{C}^T\mathsf{A}^T} = \overline{\mathsf{C}^T}\,\overline{\mathsf{A}^T} = \mathsf{C}^*\mathsf{A}^*.$$

Part (f): Recall that the trace of a matrix A is given by

$$\mathrm{tr}\,(\mathsf{A}) = A_{11} + \cdots + A_{nn} = \sum_{i=1}^{n} A_{ii}.$$

Then it is simple to see that

$$\mathrm{tr}\,(\mathsf{AB}) = \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} A_{ij}B_{ji}\Big) = \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} B_{ji}A_{ij}\Big) = \sum_{j=1}^{n}\Big(\sum_{i=1}^{m} B_{ji}A_{ij}\Big) = \mathrm{tr}\,(\mathsf{BA}).$$

This establishes the Theorem. □

We use the notation $\mathsf{A}^2 = \mathsf{AA}$, $\mathsf{A}^3 = \mathsf{A}^2\mathsf{A}$, and $\mathsf{A}^n = A^{n-1}\mathsf{A}$. Notice that the matrix product is not commutative, so the formula $(a+b)^2 = a^2 + 2ab + b^2$ does not hold for matrices. Instead we have:

$$(\mathsf{A}+\mathsf{B})^2 = (\mathsf{A}+\mathsf{B})(\mathsf{A}+\mathsf{B}) = \mathsf{A}^2 + \mathsf{AB} + \mathsf{BA} + \mathsf{B}^2.$$

2.3.4. **Block multiplication.** The multiplication of two large matrices can be simplified in the case that each matrix can be subdivided in appropriate blocks. If these matrix blocks are conformable the multiplication of the original matrices reduces to the multiplication of the smaller matrix blocks. The next result is presents a simple case.

Theorem 2.3.6. *If* A *is an* $m \times n$ *matrix and* B *is an* $n \times \ell$ *matrix having the following block decomposition,*

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{11} & \vdots & \mathsf{A}_{12} \\ \cdots & \cdots & \cdots \\ \mathsf{A}_{21} & \vdots & \mathsf{A}_{22} \end{bmatrix}, \qquad \mathsf{A} = \begin{bmatrix} m_1 \times n_1 & \vdots & m_1 \times n_2 \\ \cdots & \cdots & \cdots \\ m_2 \times n_1 & \vdots & m_2 \times n_2 \end{bmatrix},$$

$$\mathsf{B} = \begin{bmatrix} \mathsf{B}_{11} & \vdots & \mathsf{B}_{12} \\ \cdots & \cdots & \cdots \\ \mathsf{B}_{21} & \vdots & \mathsf{B}_{22} \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} n_1 \times \ell_1 & \vdots & n_1 \times \ell_2 \\ \cdots & \cdots & \cdots \\ n_2 \times \ell_1 & \vdots & n_2 \times \ell_2 \end{bmatrix},$$

where $m_1 + m_2 = m$, $n_1 + n_2 = n$ *and* $\ell_1 + \ell_2 = \ell$, *then the product* AB *has the form*

$$\mathsf{AB} = \begin{bmatrix} \mathsf{A}_{11}\mathsf{B}_{11} + \mathsf{A}_{12}\mathsf{B}_{21} & \vdots & \mathsf{A}_{11}\mathsf{B}_{12} + \mathsf{A}_{12}\mathsf{B}_{22} \\ \cdots & \cdots & \cdots \\ \mathsf{A}_{21}\mathsf{B}_{11} + \mathsf{A}_{22}\mathsf{B}_{21} & \vdots & \mathsf{A}_{21}\mathsf{B}_{12} + \mathsf{A}_{22}\mathsf{B}_{22} \end{bmatrix}, \qquad \mathsf{AB} = \begin{bmatrix} m_1 \times \ell_1 & \vdots & m_1 \times \ell_2 \\ \cdots & \cdots & \cdots \\ m_2 \times \ell_1 & \vdots & m_2 \times \ell_2 \end{bmatrix}.$$

The proof is a straightforward computation, so we omit it. This type of block decomposition is useful when several blocks are repeated inside the matrices A and B. This situation appears in the following example.

EXAMPLE 2.3.9: Use block multiplication to find the matrix AB, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 3 & 4 \end{bmatrix}.$$

SOLUTION: These matrices have the following block structure:

$$\mathsf{A} = \left[\begin{array}{cc:cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \\ \hdashline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right], \qquad \mathsf{B} = \left[\begin{array}{cc:cc} 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \\ \hdashline 0 & 0 & 1 & 2 \\ 0 & 0 & 3 & 4 \end{array}\right],$$

so, introduce the matrices

$$\mathsf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathsf{C} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathsf{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

then, the original matrices have the block form

$$\mathsf{A} = \begin{bmatrix} \mathsf{C} & \vdots & \mathsf{I} \\ \cdots & \cdots & \cdots \\ \mathsf{I} & \vdots & \mathsf{0} \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} \mathsf{I} & \vdots & \mathsf{C} \\ \cdots & \cdots & \cdots \\ \mathsf{0} & \vdots & \mathsf{C} \end{bmatrix}.$$

Then, the matrix AB has the form

$$\mathsf{AB} = \begin{bmatrix} \mathsf{C} & \vdots & \mathsf{C}^2 + \mathsf{C} \\ \cdots & \cdots & \cdots \\ \mathsf{I} & \vdots & \mathsf{C} \end{bmatrix}.$$

So, the only calculation we need to do is the matrix $\mathsf{C}^2 + \mathsf{C}$, which is given by

$$\mathsf{C}^2 + \mathsf{C} = \begin{bmatrix} 8 & 12 \\ 18 & 26 \end{bmatrix}.$$

If we put all the information above together, we get

$$\mathsf{AB} = \left[\begin{array}{cc:cc} 1 & 2 & 8 & 12 \\ 3 & 4 & 18 & 26 \\ \hdashline 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \end{array}\right] \quad \Leftrightarrow \quad \mathsf{AB} = \begin{bmatrix} 1 & 2 & 8 & 12 \\ 3 & 4 & 18 & 26 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \end{bmatrix}.$$

◁

EXAMPLE 2.3.10: Given arbitrary matrices A, that is $n \times k$, and B, that is $k \times n$, show that the $(k+n) \times (k+n)$ matrix C below satisfies $\mathsf{C}^2 = \mathsf{I}_{n+k}$, where

$$\mathsf{C} = \left[\begin{array}{c:c} \mathsf{I}_k - \mathsf{BA} & \mathsf{B} \\ \hdashline 2\mathsf{A} - \mathsf{ABA} & \mathsf{AB} - \mathsf{I}_n \end{array}\right]$$

SOLUTION: Notice that AB is an $n \times n$ matrix, while BA is an $k \times k$ matrix, so the definition of C implies that

$$\mathsf{C} = \left[\begin{array}{c:c} k \times k & k \times n \\ \hdashline n \times k & n \times n \end{array}\right].$$

Using block multiplication we obtain:

$$\mathsf{C}^2 = \left[\begin{array}{c:c} (\mathsf{I}_k - \mathsf{BA}) & \mathsf{B} \\ \hdashline (2\mathsf{A} - \mathsf{ABA}) & (\mathsf{AB} - \mathsf{I}_n) \end{array}\right] \left[\begin{array}{c:c} (\mathsf{I}_k - \mathsf{BA}) & \mathsf{B} \\ \hdashline (2\mathsf{A} - \mathsf{ABA}) & (\mathsf{AB} - \mathsf{I}_n) \end{array}\right] =$$

$$\left[\begin{array}{c:c} (\mathsf{I}_k - \mathsf{BA})(\mathsf{I}_k - \mathsf{BA}) + \mathsf{B}(2\mathsf{A} - \mathsf{ABA}) & (\mathsf{I}_k - \mathsf{BA})\mathsf{B} + \mathsf{B}(\mathsf{AB} - \mathsf{I}_n) \\ \hdashline (2\mathsf{A} - \mathsf{ABA})(\mathsf{I}_k - \mathsf{BA}) + (\mathsf{AB} - \mathsf{I}_n)(2\mathsf{A} - \mathsf{ABA}) & (2\mathsf{A} - \mathsf{ABA})\mathsf{B} + (\mathsf{AB} - \mathsf{I}_n)(\mathsf{AB} - \mathsf{I}_n) \end{array}\right]$$

Now, notice that the block $(1,1)$ above is:

$$(\mathsf{I}_k - \mathsf{BA})(\mathsf{I}_k - \mathsf{BA}) + \mathsf{B}(2\mathsf{A} - \mathsf{ABA}) = \mathsf{I}_k - \mathsf{BA} - \mathsf{BA} + \mathsf{BABA} + 2\mathsf{BA} - \mathsf{BABA} = \mathsf{I}_k.$$

The block $(1,2)$ is:

$$(\mathsf{I}_k - \mathsf{BA})\mathsf{B} + \mathsf{B}(\mathsf{AB} - \mathsf{I}_n) = \mathsf{B} - \mathsf{BAB} + \mathsf{BAB} - \mathsf{B} = 0.$$

The block $(2,1)$ is:

$$(2\mathsf{A} - \mathsf{ABA})(\mathsf{I}_k - \mathsf{BA}) + (\mathsf{AB} - \mathsf{I}_n)(2\mathsf{A} - \mathsf{ABA}) =$$

$$2\mathsf{A} - 2\mathsf{ABA} - \mathsf{ABA} + \mathsf{ABABA} + 2\mathsf{ABA} - \mathsf{ABABA} - 2\mathsf{A} + \mathsf{ABA} = 0.$$

Finally, the block $(2,2)$ is:

$$(2\mathsf{A} - \mathsf{ABA})\mathsf{B} + (\mathsf{AB} - \mathsf{I}_n)(\mathsf{AB} - \mathsf{I}_n) = 2\mathsf{AB} - \mathsf{ABAB} + \mathsf{ABAB} - \mathsf{AB} - \mathsf{AB} + \mathsf{I}_n = \mathsf{I}_n.$$

Therefore, we have shown that: $\mathsf{C}^2 = \begin{bmatrix} \mathsf{I}_k & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \mathsf{I}_n \end{bmatrix} = \mathsf{I}_{k+n}$. ◁

2.3.5. **Matrix commutators.** Matrix multiplication is in general not commutative. Given two $n \times n$ matrices A and B, their product AB is in general different from BA. We call the difference between these two matrices the commutator of A and B.

Definition 2.3.7. *The* ***commutator*** *of two $n \times n$ matrices A and B, is the matrix*

$$[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}.$$

Furthermore, we say that the square matrices A, B ***commute*** *iff holds $[\mathsf{A}, \mathsf{B}] = 0$.*

Therefore, the symbol $[\mathsf{A}, \mathsf{B}]$ denotes an $n \times n$ matrix, the commutator of A and B. The commutator of two operators defined on an inner product space is an important concept in quantum mechanics. When we say an operator, we should think in something like a matrix, and when we say an inner product space, we should think in something like $\mathbb{F}^n$. An observation of a physical property, like position or momentum, of a quantum mechanical system, like an electron, is described by an operator acting on a vector, the latter describing the state of the quantum mechanical system. By observing nature, one discovers that two properties can be simultaneously measured without limit in the measurement precision iff the associated operators commute. If the operators do not commute, then their commutator determines the maximum precision of their simultaneous observation. This is the physical content of the *Heisenberg uncertainty principle*.

EXAMPLE 2.3.11: We have seen in Example 2.3.6 that two rotation matrices $\mathsf{R}(\theta_1)$ and $\mathsf{R}(\theta_2)$ commute for all $\theta_1, \theta_2 \in \mathbb{R}$, that is,

$$[\mathsf{R}(\theta_1), \mathsf{R}(\theta_2)] = \mathsf{R}(\theta_1)\mathsf{R}(\theta_2) - \mathsf{R}(\theta_2)\mathsf{R}(\theta_1) \quad \Rightarrow \quad [\mathsf{R}(\theta_1), \mathsf{R}(\theta_2)] = 0.$$ ◁

Theorem 2.3.8. *For all matrices A, B, $\mathsf{C} \in \mathbb{F}^{n,n}$ and scalars a, b, $c \in \mathbb{F}$ holds:*

(a) $[\mathsf{A}, \mathsf{B}] = -[\mathsf{B}, \mathsf{A}]$, *(antisymmetry);*
(b) $[a\mathsf{A}, b\mathsf{B}] = ab\,[\mathsf{A}, \mathsf{B}]$, *(linearity);*
(c) $[\mathsf{A}, \mathsf{B} + \mathsf{C}] = [\mathsf{A}, \mathsf{B}] + [\mathsf{A}, \mathsf{C}]$, *(linearity on the right entry);*
(d) $[\mathsf{A} + \mathsf{B}, \mathsf{C}] = [\mathsf{A}, \mathsf{C}] + [\mathsf{B}, \mathsf{C}]$, *(linearity on the left entry);*
(e) $[\mathsf{A}, \mathsf{BC}] = [\mathsf{A}, \mathsf{B}]\mathsf{C} + \mathsf{B}[\mathsf{A}, \mathsf{C}]$, *(right derivation property);*
(f) $[\mathsf{AB}, \mathsf{C}] = [\mathsf{A}, \mathsf{C}]\mathsf{B} + \mathsf{A}[\mathsf{B}, \mathsf{C}]$, *(left derivation property);*
(g) $\big[[\mathsf{A}, \mathsf{B}], \mathsf{C}\big] + \big[[\mathsf{C}, \mathsf{A}], \mathsf{B}\big] + \big[[\mathsf{B}, \mathsf{C}], \mathsf{A}\big] = 0$, *(Jacobi property).*

Proof of Theorem 2.3.8: All properties are simple to show.

Part (a):

$$[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA} = -\big(\mathsf{BA} - \mathsf{AB}\big) = -[\mathsf{B}, \mathsf{A}].$$

Part (b):

$$[a\mathsf{A}, b\mathsf{B}] = a\mathsf{A}b\mathsf{B} - b\mathsf{B}a\mathsf{A} = ab\big(\mathsf{AB} - \mathsf{BA}\big) = ab\,[\mathsf{A}, \mathsf{B}].$$

Part (c):

$$[\mathsf{A}, \mathsf{B} + \mathsf{C}] = \mathsf{A}(\mathsf{B} + \mathsf{C}) - (\mathsf{B} + \mathsf{C})\mathsf{A} = \mathsf{AB} + \mathsf{AC} - \mathsf{BA} - \mathsf{CA} = [\mathsf{A}, \mathsf{B}] + [\mathsf{A}, \mathsf{C}].$$

Part (d): The proof is similar to the previous calculation.

Part (e):

$$[\mathsf{A}, \mathsf{BC}] = \mathsf{A}(\mathsf{BC}) - (\mathsf{BC})\mathsf{A} = \mathsf{ABC} + (\mathsf{BAC} - \mathsf{BAC}) - \mathsf{BCA} = [\mathsf{A}, \mathsf{B}]\mathsf{C} + \mathsf{B}[\mathsf{A}, \mathsf{C}].$$

Part (f): The proof is similar to the previous calculation.

Part (g): We write down each term and we verify that they all add up to zero:

$$\big[[\mathsf{A},\mathsf{B}],\mathsf{C}\big] = (\mathsf{AB}-\mathsf{BA})\mathsf{C}-\mathsf{C}(\mathsf{AB}-\mathsf{BA});$$
$$\big[[\mathsf{C},\mathsf{A}],\mathsf{B}\big] = (\mathsf{CA}-\mathsf{AC})\mathsf{B}-\mathsf{B}(\mathsf{CA}-\mathsf{AC});$$
$$\big[[\mathsf{B},\mathsf{C}],\mathsf{A}\big] = (\mathsf{BC}-\mathsf{CB})\mathsf{A}-\mathsf{A}(\mathsf{BC}-\mathsf{CB}).$$

These three equations add up to the zero matrix. This establishes the Theorem. □

We finally highlight that these properties implies that $[\mathsf{A},\mathsf{A}^m] = 0$ holds for all $m \in \mathbb{N}$.

Further reading. Almost every book in linear algebra explains matrix multiplication. See Section 3.5 in Meyer's book [3], and Section 3.6 for block multiplication. Also Strang's book [4]. The definition of commutators can be found in Section 2.3.1 in Hassani's book [1]. Commutators play an important role in the Spectral Theorem, which we study in Chapter 9.

2.3.6. **Exercises.**

2.3.1.- Given the matrices

$$\mathsf{A} = \begin{bmatrix} 1 & -2 & 3 \\ 0 & -5 & 4 \\ 4 & -3 & 8 \end{bmatrix}, \ \mathsf{B} = \begin{bmatrix} 1 & 2 \\ 0 & 4 \\ 3 & 7 \end{bmatrix},$$

and the vector $\mathsf{C} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, compute the following products, if possible:

(a) AB, BA, CB and $\mathsf{C}^T\mathsf{B}$.

(b) A^2, B^2, $\mathsf{C}^T\mathsf{C}$ and $\mathsf{C}\mathsf{C}^T$.

2.3.2.- Consider the matrices

$$\mathsf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix}, \mathsf{B} = \begin{bmatrix} 1 & 1 \\ 3 & 0 \end{bmatrix}, \mathsf{C} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}.$$

(a) Compute $[\mathsf{A}, \mathsf{B}]$.

(b) Find the product ABC.

2.3.3.- Find A^2 and A^3 for the matrix

$$\mathsf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

2.3.4.- Given a real number a find the matrix A^n, where n is any positive integer and

$$\mathsf{A} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}.$$

2.3.5.- Given any square matrices A, B, prove that

$$(\mathsf{A}+\mathsf{B})^2 = \mathsf{A}^2+2\mathsf{AB}+\mathsf{B}^2 \ \Leftrightarrow \ [\mathsf{A}, \mathsf{B}] = 0.$$

2.3.6.- Given $a = 1/3$, divide the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & 0 & a & a & a \\ 0 & 1 & 0 & a & a & a \\ 0 & 0 & 1 & a & a & a \\ 0 & 0 & 0 & a & a & a \\ 0 & 0 & 0 & a & a & a \\ 0 & 0 & 0 & a & a & a \end{bmatrix}$$

into appropriate blocks, and using block multiplication find the matrix A^{300}.

2.3.7.- Prove that for all matrices $\mathsf{A} \in \mathbb{F}^{n,m}$ and $\mathsf{B} \in \mathbb{F}^{m,n}$ holds

$$\mathrm{tr}\,(\mathsf{AB}) = \mathrm{tr}\,(\mathsf{BA}).$$

2.3.8.- Let A be an $m \times n$ matrix. Show that $\mathrm{tr}\left(\mathsf{A}^T\mathsf{A}\right) = 0$ iff $\mathsf{A} = \mathsf{0}$.

2.3.9.- Prove that: If A, $\mathsf{B} \in \mathbb{F}^{n,n}$ are symmetric matrices and commute, then their product AB is also a symmetric matrix.

2.3.10.- Let A be an arbitrary $n \times n$ matrix. Use the trace function to show that there exists no $n \times n$ matrix X solution of the matrix equation

$$[\mathsf{A}, \mathsf{X}] = \mathsf{I}_n.$$

2.4. Inverse matrix

In this Section we introduce the concept of the inverse of a square matrix. Not every square matrix is invertible. In the case of 2×2 matrices we present a condition on the matrix coefficients that is equivalent to the invertibility of the matrix, and we also present a formula for the inverse matrix. Later on in Section 3.2 we generalize this formula for $n \times n$ matrices. The inverse of a matrix is useful to compute solutions to systems of linear equations.

2.4.1. **Main definition.** We start recalling that the matrix $\mathsf{I}_n \in \mathbb{F}^{n,n}$ is called the *identity matrix* iff holds that $\mathsf{I}_n \mathsf{x} = \mathsf{x}$ for all $\mathsf{x} \in \mathbb{F}^n$. Choosing the vector $\mathsf{x} = \mathsf{e}_i$, consisting of a 1 in the row i and zero every where else, it is simple to see that the components of the identity matrix are given by

$$\mathsf{I}_n = [I_{ij}] \quad \text{with} \quad \begin{cases} I_{ii} = 1, \\ I_{ij} = 0, \quad i \neq j. \end{cases}$$

The cases $n = 2$ and $n = 3$ are shown below,

$$\mathsf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathsf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We also use the column vector notation for the identity matrix,

$$\mathsf{I}_n = [\mathsf{e}_1, \cdots, \mathsf{e}_n],$$

that is, we denote the identity column vectors $\mathsf{I}_{:i} = \mathsf{e}_i$ for $i = 1, \cdots, n$. For example, in the case of I_3 we have that

$$\mathsf{I}_3 = [\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathsf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We are now ready to introduce the notion of the inverse matrix.

Definition 2.4.1. *A matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ is called **invertible** iff there exists a matrix, denoted as A^{-1}, such that $(\mathsf{A}^{-1})\mathsf{A} = \mathsf{I}_n$ and $\mathsf{A}(\mathsf{A}^{-1}) = \mathsf{I}_n$.*

Since the matrix product is non-commutative, the products $\mathsf{A}(\mathsf{A}^{-1}) = \mathsf{I}_n$ and $(\mathsf{A}^{-1})\mathsf{A} = \mathsf{I}_n$ must be specified in the definition above. Notice that we do not need to assume that the inverse matrix belongs to $\mathbb{F}^{n,n}$, since both products $\mathsf{A}(\mathsf{A}^{-1})$ and $(\mathsf{A}^{-1})\mathsf{A}$ are well-defined, we conclude that the inverse matrix must be $n \times n$.

Example 2.4.1: Verify that the matrix and its inverse are given by

$$\mathsf{A} = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \mathsf{A}^{-1} = \frac{1}{4} \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}.$$

Solution: We have to compute the products,

$$\mathsf{A}(\mathsf{A}^{-1}) = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad \Rightarrow \quad \mathsf{A}(\mathsf{A}^{-1}) = \mathsf{I}_2.$$

It is simple to check that the equation $(\mathsf{A}^{-1})\mathsf{A} = \mathsf{I}_2$ also holds. ◁

Example 2.4.2: The only real numbers that are equal to is own inverses are $a = 1$ and $a = -1$. This is not true in the case of matrices. Verify that the matrix A below is its own inverse, that is,

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \mathsf{A}^{-1}.$$

SOLUTION: We have to compute the products,

$$A(A^{-1}) = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Rightarrow \quad A(A^{-1}) = I_2.$$

It is simple to check that the equation $(A^{-1})A = I_2$ also holds. ◁

Not every square matrix is invertible. The following Example show a 2×2 matrix with no inverse.

EXAMPLE 2.4.3: Show that matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ has no inverse.

SOLUTION: Suppose there exists the inverse matrix

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then, the following equation holds,

$$A(A^{-1}) = I_2 \quad \Leftrightarrow \quad \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The last equation implies

$$a + 2c = 1, \qquad b + 2d = 0,$$
$$3(a + 2c) = 0, \qquad 3(b + 2d) = 1.$$

However, both systems are inconsistent, so the inverse matrix A^{-1} does not exist. ◁

In the case of 2×2 matrices there is a simple way to find out whether a matrix has inverse or not. If the 2×2 matrix is invertible, then there is a simple formula for the inverse matrix. This is summarized in the following result.

Theorem 2.4.2. *Given a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, introduce the number $\Delta = ad - bc$. The matrix A is invertible iff $\Delta \neq 0$. Furthermore, if A is invertible, its inverse is given by*

$$A^{-1} = \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{2.5}$$

The number Δ is called the *determinant of* A, since it is the number that determines whether A is invertible or not, and soon we will see that it is the number that determines whether a system of linear equations has a unique solution or not. Also later on we will study generalizations to $n \times n$ matrices of the Theorem above. That will require a generalization to $n \times n$ matrices of the determinant Δ of a matrix.

EXAMPLE 2.4.4: Compute the inverse of matrix $A = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$, given in Example 2.4.1.

SOLUTION: Following Theorem 2.4.2 we first compute $\Delta = 6 - 4 = 4$. Since $\Delta \neq 0$, then A^{-1} exists and it is given by

$$A^{-1} = \frac{1}{4} \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}.$$

◁

EXAMPLE 2.4.5: Theorem 2.4.2 says that the matrix in Example 2.4.3 is not invertible, since

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \quad \Rightarrow \quad \Delta = 6 - (3)(2) \quad \Rightarrow \quad \Delta = 0.$$

◁

Proof of Theorem 2.4.2: If the matrix A^{-1} exists, from the definition of inverse matrix it follows that A^{-1} must be 2×2. Suppose that the inverse of matrix A is given by

$$\mathsf{A}^{-1} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}.$$

We first show that A^{-1} exists iff $\Delta \neq 0$.

The 2×2 matrix A^{-1} exists iff the equation $\mathsf{A}(\mathsf{A}^{-1}) = \mathsf{I}_2$, which is equivalent to the systems

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Leftrightarrow \quad \begin{array}{ll} ax_1 + bx_2 = 1, & ay_1 + by_2 = 0, \\ cx_1 + dx_2 = 0, & cy_1 + dy_2 = 1. \end{array} \tag{2.6}$$

Consider now the particular case $a = 0$ and $c = 0$, which imply that $\Delta = 0$. In this case the equations above reduce to:

$$\begin{aligned} bx_2 &= 1, & by_2 &= 0, \\ dx_2 &= 0, & dy_2 &= 1. \end{aligned}$$

These systems have no solution, since from the first equation on the left $b \neq 0$, and from the first equation on the right we obtain that $y_2 = 0$. But this contradicts the second equation on the right, since dy_2 is zero and so can never be equal to one. We then conclude there is no inverse matrix in this case.

Assume now that at least one of the coefficients a or c is non-zero, and let us return to Eqs. (2.6). Both systems can be solved using the following augmented matrix

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right]$$

Now, perform the following Gauss operations:

$$\left[\begin{array}{cc|cc} ac & bc & c & 0 \\ ac & ad & 0 & a \end{array}\right] \to \left[\begin{array}{cc|cc} ac & bc & c & 0 \\ 0 & ad - bc & -c & a \end{array}\right] = \left[\begin{array}{cc|cc} ac & bc & c & 0 \\ 0 & \Delta & -c & a \end{array}\right]$$

At least one of the source coefficients in the second row above is non-zero. Therefore, the system above is consistent iff $\Delta \neq 0$. This establishes the first part of the Theorem.

In order to prove the furthermore part, one can continue with the calculation above, and find the formula for the inverse matrix. This is a long calculation, since one has to study three different cases: the case $a = 0$ and $c \neq 0$, the case $a \neq 0$ and $c = 0$, and the case where both a and c are non-zero. It is faster to check that the expression in the Theorem is indeed the inverse of A. Since $\Delta \neq 0$, the matrix in Eq. (2.5) is well-defined. Then, the straightforward calculation

$$\mathsf{A}(\mathsf{A}^{-1}) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \Delta & -ab + ba \\ cd - dc & \Delta \end{bmatrix} = \mathsf{I}_2.$$

It is not difficult to see that the second condition $(\mathsf{A}^{-1})\mathsf{A} = \mathsf{I}_2$ is also satisfied. This establishes the Theorem. □

There are many different ways to characterize $n \times n$ invertible matrices. One possibility is to relate the existence of the inverse matrix to the solution of appropriate systems of linear equations. The following result summarizes a few of these characterizations.

Theorem 2.4.3. *Given a matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, the following statements are equivalent:*

(a) The matrix A^{-1} exists;
(b) $\text{rank}(\mathsf{A}) = n$;
(c) $\mathsf{E}_\mathsf{A} = \mathsf{I}_n$;
(d) The homogeneous equation $\mathsf{A}\mathsf{x} = \mathsf{0}$ has a unique solution $\mathsf{x} = \mathsf{0}$;
(e) The non-homogeneous equation $\mathsf{A}\mathsf{x} = \mathsf{b}$ has a unique solution for every source $\mathsf{b} \in \mathbb{F}^n$.

Proof of Theorem 2.4.3: It is clear that properties (b)-(d) are all equivalent. Here we only show that (a) is equivalent to (b).

We start with (a) ⇒ (b): Since A^{-1} exists, the homogeneous equation $\mathsf{Ax} = \mathsf{0}$ has a unique solution $\mathsf{x} = \mathsf{0}$. (Proof: Assume there are two solutions x_1 and x_2, then $\mathsf{A}(\mathsf{x}_2 - \mathsf{x}_1) = \mathsf{0}$, and so $(\mathsf{x}_2 - \mathsf{x}_1) = \mathsf{A}^{-1}\mathsf{A}(\mathsf{x}_2 - \mathsf{x}_1) = \mathsf{A}^{-1}\mathsf{0} = \mathsf{0}$, and so $\mathsf{x}_2 = \mathsf{x}_1$.) But this implies there are no free variables in the solutions of $\mathsf{Ax} = \mathsf{0}$, that is, $\mathsf{E}_\mathsf{A} = \mathsf{I}_n$, that is, $\text{rank}(\mathsf{A}) = n$.

We finish with (b) ⇒ (a): Since $\text{rank}(\mathsf{A}) = n$, the non-homogeneous equation $\mathsf{Ax} = \mathsf{b}$ has a unique solution x for every source vector $\mathsf{b} \in \mathbb{R}^n$. In particular, there exist unique vectors $\mathsf{x}_1, \cdots, \mathsf{x}_n$ solutions of

$$\mathsf{Ax}_1 = \mathsf{e}_1, \quad \cdots \quad \mathsf{Ax}_n = \mathsf{e}_n.$$

where e_i is the i-th column of the identity matrix I_n. This is equivalent to say that the matrix $\mathsf{X} = \big[\mathsf{x}_1, \cdots, \mathsf{x}_n\big]$ satisfies the equation

$$\mathsf{AX} = \mathsf{I}_n.$$

This matrix $\mathsf{X} = \mathsf{A}^{-1}$ since X also satisfies the equation $\mathsf{XA} = \mathsf{I}_n$. (Proof: Consider the identities

$$\mathsf{A} - \mathsf{A} = \mathsf{0} \quad \Leftrightarrow \quad \mathsf{AXA} - \mathsf{A} = \mathsf{0} \quad \Leftrightarrow \quad \mathsf{A}(\mathsf{XA} - \mathsf{I}_n) = \mathsf{0};$$

The last equation is equivalent to the n systems of equations $\mathsf{Ay}_i = \mathsf{0}$, where y_i is the i-th column of the matrix $(\mathsf{XA} - \mathsf{I}_n)$; Since $\text{rank}(\mathsf{A}) = n$, each of these systems has a unique solution $\mathsf{y}_i = \mathsf{0}$, that is, $\mathsf{XA} = \mathsf{I}_n$.) This establishes the Theorem. □

It is simple to see from Theorem 2.4.3 that an invertible matrix has a unique inverse.

Corollary 2.4.4. *An invertible matrix has a unique inverse.*

Proof of Corollary 2.4.4: Suppose that matrices X and Y are two inverses of an $n \times n$ matrix A. Then, $\mathsf{AX} = \mathsf{I}_n$ and $\mathsf{AY} = \mathsf{I}_n$, hence $\mathsf{A}(\mathsf{X} - \mathsf{Y}) = \mathsf{0}$. The latter are n systems of linear equations, one for each column vector in $(\mathsf{X} - \mathsf{Y})$. From Theorem 2.4.3 we know that $\text{rank}(\mathsf{A}) = n$, so the only solution to these equations is the trivial solution, so each column vector vanishes, therefore $\mathsf{X} = \mathsf{Y}$. □

2.4.2. **Properties of invertible matrices.** They are summarized below.

Theorem 2.4.5. *If* A *and* B *are* $n \times n$ *invertible matrices, then holds:*

(a) $\big(\mathsf{A}^{-1}\big)^{-1} = \mathsf{A}$;

(b) $\big(\mathsf{AB}\big)^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}$;

(c) $\big(\mathsf{A}^T\big)^{-1} = \big(\mathsf{A}^{-1}\big)^T$.

Proof of Theorem 2.4.5: Since an invertible matrix is unique, we only have to verify these equations in (a)-(c).

Part (a): The inverse of A^{-1} is a matrix $\big(\mathsf{A}^{-1}\big)^{-1}$ satisfying the equations

$$\Big[\big(\mathsf{A}^{-1}\big)^{-1}\Big]\big(\mathsf{A}^{-1}\big) = \mathsf{I}_n, \qquad \big(\mathsf{A}^{-1}\big)\Big[\big(\mathsf{A}^{-1}\big)^{-1}\Big] = \mathsf{I}_n.$$

But matrix A satisfies precisely these equations, and recalling that the inverse of a matrix is unique, then $\mathsf{A} = \big(\mathsf{A}^{-1}\big)^{-1}$.

Part (b): The proof is similar to the previous one. We verify that matrix $\big(\mathsf{B}^{-1}\big)\big(\big(\mathsf{A}^{-1}\big)$ satisfies the equations that $(\mathsf{AB})^{-1}$ must satisfy. Then, they must be the same, since the inverse matrix is unique. Notice that,

$$\begin{aligned} \big(\mathsf{B}^{-1}\mathsf{A}^{-1}\big)(\mathsf{AB}) &= \mathsf{B}^{-1}\big(\mathsf{A}^{-1}\mathsf{A}\big)\mathsf{B}, \\ &= \big(\mathsf{B}^{-1}\big)\mathsf{B}, \\ &= \mathsf{I}_n; \end{aligned} \qquad \begin{aligned} (\mathsf{AB})\big(\mathsf{B}^{-1}\mathsf{A}^{-1}\big) &= \mathsf{A}\big(\mathsf{B}\mathsf{B}^{-1}\big)\mathsf{A}^{-1}, \\ &= \mathsf{A}\big(\mathsf{A}^{-1}\big), \\ &= \mathsf{I}_n. \end{aligned}$$

We then conclude that $\big(AB\big)^{-1} = B^{-1}A^{-1}$.

Part (c): Recall that $(AB)^T = B^T A^T$, therefore,

$$\big(A^{-1}\big)A = I_n \quad \Leftrightarrow \quad \big[\big(A^{-1}\big)A\big]^T = I_n^T \quad \Leftrightarrow \quad A^T\big(A^{-1}\big)^T = I_n,$$

$$A\big(A^{-1}\big) = I_n \quad \Leftrightarrow \quad \big[A\big(A^{-1}\big)\big]^T = I_n^T \quad \Leftrightarrow \quad \big(A^{-1}\big)^T A^T = I_n.$$

Therefore, $\big(A^{-1}\big)^T = \big(A^T\big)^{-1}$. This establishes the Theorem. □

The properties presented above are useful to solve equations involving matrices.

EXAMPLE 2.4.6: Find a matrix C solution of the equation $(BC)^T - A = 0$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -1 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}.$$

SOLUTION: The matrix B is invertible, since $\Delta(B) = 3$, and its inverse is given by

$$B^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}.$$

Therefore, matrix C is given by

$$(BC)^T = A \quad \Leftrightarrow \quad BC = A^T \quad \Leftrightarrow \quad C = B^{-1}A^T,$$

that is,

$$C = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 & -2 \\ 2 & 4 & -1 \end{bmatrix} \quad \Rightarrow \quad C = \frac{1}{3}\begin{bmatrix} 4 & 10 & -5 \\ 1 & 1 & 1 \end{bmatrix}.$$

◁

EXAMPLE 2.4.7: The $(n+k)\times(n+k)$ matrix C introduced in Example 2.3.10 satisfies the equation $C^2 = I_{n+k}$. Therefore, matrix C is its own inverse, that is, $C^{-1} = C$. ◁

2.4.3. **Computing the inverse matrix.** We now show how to use Gauss operations to find the inverse matrix in the case that such inverse exists. The main idea needed to compute the inverse of an $n \times n$ matrix is summarized in the following Theorem. We emphasize that this is not a new result. We are just highlighting the main part of the proof in Theorem 2.4.3, (b) ⇒ (a).

Theorem 2.4.6. *Let A be an $n \times n$ matrix. If the n systems of linear equations*

$$Ax_1 = e_1, \quad \cdots \quad Ax_n = e_n, \tag{2.7}$$

are all consistent, then the matrix A is invertible and its inverse is given by

$$A^{-1} = \big[x_1, \cdots, x_n\big].$$

If at least one system in Eq. (2.7) is inconsistent, then matrix A is not invertible.

Proof of Theorem 2.4.6: We first show that the consistency of all systems in Eq. (2.7) implies that matrix A is invertible. Indeed, if all systems in Eq. (2.7) are consistent, then the system $Ax = b$ is also consistent for all $b \in \mathbb{R}^n$ (Proof: The solution for the source $b = b_1 e_1 + \cdots + b_n e_n$ is simply $x = b_1 x_1 + \cdots + b_n x_n$.) Therefore, $\text{rank}(A) \geqslant n$. Since A is an $n \times n$ matrix, we conclude that $\text{rank}(A) = n$. Then, Theorem 2.4.3 implies that matrix A is invertible.

We now introduce the matrix $X = \big[x_1, \cdots, x_n\big]$ and we show that $A^{-1} = X$. From the definition of X we see that $AX = I_n$. Since $\text{rank}(A) = n$, the same argument given in the proof of Theorem 2.4.3 shows that $XA = I_n$. Since the inverse of a matrix is unique, we conclude that $A^{-1} = X$. This establishes the Theorem. □

This Theorem provides a method to find the inverse of a matrix.

Corollary 2.4.7. *The inverse of an invertible $n \times n$ matrix A is computed with the Gauss operations such that $\big[\mathsf{A}\big|\mathsf{I}_n\big] \longrightarrow \big[\mathsf{I}_n\big|\mathsf{A}^{-1}\big]$.*

Proof of Corollary 2.4.7: One solves the n linear systems in Eq. (2.7). Since all these systems share the same coefficient matrix, matrix A, one can solve them all at the same time, introducing the augmented matrix

$$\big[\mathsf{A}\big|\mathsf{e}_1, \cdots, \mathsf{e}_n\big] = \big[\mathsf{A}\big|\mathsf{I}_n\big].$$

In the proof of Theorem 2.4.6 we show that $\text{rank}(\mathsf{A}) = n$. Therefore, its reduced echelon form is $\mathsf{E}_\mathsf{A} = \mathsf{I}_n$, so the Gauss method implies that

$$\big[\mathsf{A}\big|\mathsf{I}_n\big] = \big[\mathsf{A}\big|\mathsf{e}_1, \cdots, \mathsf{e}_n\big] \to \big[\mathsf{E}_\mathsf{A}\big|\mathsf{x}_1, \cdots, \mathsf{x}_n\big] = \big[\mathsf{I}_n\big|\mathsf{A}^{-1}\big] \quad \Leftrightarrow \quad \big[\mathsf{A}\big|\mathsf{I}_n\big] \longrightarrow \big[\mathsf{I}_n\big|\mathsf{A}^{-1}\big].$$

This establishes the Corollary. □

EXAMPLE 2.4.8: Find the inverse of the matrix $\mathsf{A} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$.

SOLUTION: We have to solve the two systems of equations $\mathsf{A}\mathsf{x}_1 = \mathsf{e}_1$ and $\mathsf{A}\mathsf{x}_2 = \mathsf{e}_2$, that is,

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We solve both systems at the same time using the Gauss method on the augmented matrix

$$\left[\begin{array}{cc|cc} 2 & 4 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1/2 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1/2 & 0 \\ 0 & 1 & -1/2 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 0 & 3/2 & -2 \\ 0 & 1 & -1/2 & 1 \end{array}\right]$$

therefore, $\mathsf{A}^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix}$. ◁

EXAMPLE 2.4.9: Find the inverse of the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 9 \end{bmatrix}$.

SOLUTION: We compute the inverse matrix as follows:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 7 & 0 & 1 & 0 \\ 3 & 7 & 9 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -3 & 0 & 1 \end{array}\right] \to$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 5 & -2 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 0 & -1 & -1 & -1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & -3 & 1 \\ 0 & 1 & 0 & -3 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right] \Rightarrow \mathsf{A}^{-1} = \begin{bmatrix} 4 & -3 & 1 \\ -3 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

◁

EXAMPLE 2.4.10: Is matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ invertible?

SOLUTION: No, since $\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array}\right]$, are inconsistent. ◁

Further reading. Almost any book in linear algebra introduces the inverse matrix. See Sections 3.7 and 3.9 in Meyer's book [3].

2.4.4. **Exercises.**

2.4.1.- When possible, find the inverse of the following matrices. (Check your answers using matrix multiplication.)

(a)
$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix};$$

(b)
$$\mathsf{A} = \begin{bmatrix} -1 & -2 & 1 \\ 3 & 2 & -6 \\ 1 & 1 & -2 \end{bmatrix};$$

(c)
$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

2.4.2.- Find the values of the constant k such that the matrix A below is not invertible,
$$\mathsf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & k \\ 1 & k & 3 \end{bmatrix}.$$

2.4.3.- Consider the matrix
$$\mathsf{A} = \begin{bmatrix} 0 & -1 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix},$$

(a) Find the inverse of matrix A.

(b) Use the previous part to find a solution to the linear system
$$\mathsf{A}\mathsf{x} = \mathsf{b}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}.$$

2.4.4.- Show that for every invertible matrix A holds that $[\mathsf{A}, \mathsf{A}^{-1}] = \mathsf{0}$.

2.4.5.- Find a matrix X such that the equation $\mathsf{X} = \mathsf{A}\mathsf{X} + \mathsf{B}$ holds for
$$\mathsf{A} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}.$$

2.4.6.- If A is invertible and symmetric, then show that A^{-1} is also symmetric.

2.4.7.- Prove that: If the square matrix A satisfies $\mathsf{A}^2 = \mathsf{0}$, then the matrix $(\mathsf{I} - \mathsf{A})$ is invertible.

2.4.8.- Prove that: If the square matrix A satisfies $\mathsf{A}^3 = \mathsf{0}$, then the matrix $(\mathsf{I} - \mathsf{A})$ is invertible.

2.4.9.- Let A be a square matrix. Prove the following statements:

(a) If A contains a zero column then A is not invertible;

(b) If one column is multiple of another column in A, then matrix A is not invertible.

(c) Use the trace function to prove the following statement: If A is an $m \times n$ matrix and B is an $n \times m$ matrix such that $\mathsf{A}\mathsf{B} = \mathsf{I}_m$ and $\mathsf{B}\mathsf{A} = \mathsf{I}_n$, then $m = n$.

2.4.10.- Consider the invertible matrices $\mathsf{A} \in \mathbb{F}^{r,r}$, $\mathsf{B} \in \mathbb{F}^{s,s}$ and the matrix $\mathsf{C} \in \mathbb{F}^{r,s}$. Prove that the inverse of
$$\left[\begin{array}{c:c} \mathsf{A} & \mathsf{C} \\ \hdashline \mathsf{0} & \mathsf{B} \end{array}\right]$$
is given by
$$\left[\begin{array}{c:c} \mathsf{A}^{-1} & -\mathsf{A}^{-1}\mathsf{C}\,\mathsf{B}^{-1} \\ \hdashline \mathsf{0} & \mathsf{B}^{-1} \end{array}\right].$$

2.5. Null and range spaces

2.5.1. **Definition of the spaces.** A matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ defines two functions, $\mathsf{A} : \mathbb{F}^n \to \mathbb{F}^m$ and $\mathsf{A}^T : \mathbb{F}^m \to \mathbb{F}^n$, which in turn, determine two sets in $\mathbb{F}^n$ and two sets in $\mathbb{F}^m$. In this Section we define these sets, we call them null and range spaces, and study the main relations among them. We start introducing the two spaces associated with the function A.

Definition 2.5.1. *Consider the matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ defining the linear function $\mathsf{A} : \mathbb{F}^n \to \mathbb{F}^m$. The **null space** of the function A is the set $N(\mathsf{A}) \subset \mathbb{F}^n$ given by*

$$N(\mathsf{A}) = \{\mathsf{x} \in \mathbb{F}^n \ : \ \mathsf{Ax} = \mathsf{0}\}.$$

*The **range space** of the function A is set $R(\mathsf{A}) \subset \mathbb{F}^m$ given by*

$$R(\mathsf{A}) = \{\mathsf{y} \in \mathbb{F}^m \ : \ \mathsf{y} = \mathsf{Ax}, \ \textit{for all } \mathsf{x} \in \mathbb{F}^n\}.$$

In Fig. 25 we show a picture, usual in set theory, sketching the null and range spaces associated with a matrix A. One can see in these pictures that for a function $\mathsf{A} : \mathbb{F}^n \to \mathbb{F}^m$, the null space is a subset in $\mathbb{F}^n$, while the range space is a subset in $\mathbb{F}^m$. Recall that the homogeneous equation $\mathsf{Ax} = \mathsf{0}$ always has the trivial solution $\mathsf{x} = \mathsf{0}$. This property implies both that $\mathsf{0} \in \mathbb{F}^n$ also belongs to $N(\mathsf{A})$ and that $\mathsf{0} \in \mathbb{F}^m$ also belongs to $R(\mathsf{A})$.

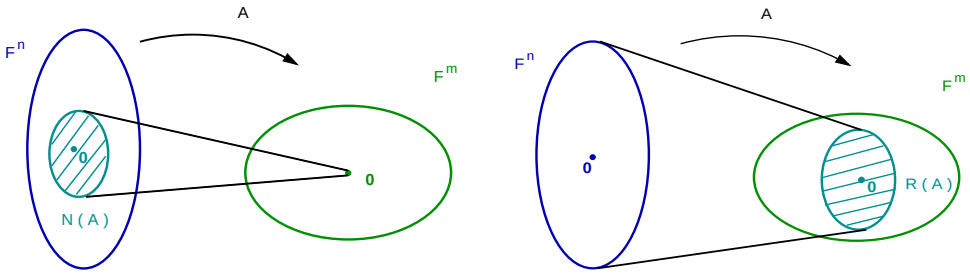

FIGURE 25. Sketch of the null space, $N(\mathsf{A})$, and the range space, $R(\mathsf{A})$, of a linear function determined by the $m \times n$ matrix A.

The *four spaces* associated with a matrix A are $N(\mathsf{A})$ and $R(\mathsf{A})$ together with $N(\mathsf{A}^T)$ and $R(\mathsf{A}^T)$. Notice that these null and range spaces are subsets of different spaces. More precisely, a matrix $\mathsf{A} \in \mathbb{F}^{m.n}$ defines the functions and subsets,

$$\mathsf{A} : \mathbb{F}^n \to \mathbb{F}^m \quad \text{and} \quad \mathsf{A}^T : \mathbb{F}^m \to \mathbb{F}^n,$$

$$N(\mathsf{A}), R(\mathsf{A}^T) \subset \mathbb{F}^n, \quad \text{while} \quad R(\mathsf{A}), N(\mathsf{A}^T) \subset \mathbb{F}^m.$$

EXAMPLE 2.5.1: Find the $N(\mathsf{A})$ and $R(\mathsf{A})$ for the function $\mathsf{A} : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}.$$

SOLUTION: We first find $N(\mathsf{A})$, the null space of A. This is the set of elements $\mathsf{x} \in \mathbb{R}^3$ that are solutions of the equation $\mathsf{Ax} = \mathsf{0}$. We use the Gauss method to find all such solutions,

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & -5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = -2x_2 \\ x_3 = 0, \\ x_2 : \text{free variable.} \end{cases}$$

Therefore, the elements in $N(\mathsf{A})$ are given by

$$N(\mathsf{A}) \ni \mathsf{x} = \begin{bmatrix} -2x_2 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} x_2 \quad \Rightarrow \quad N(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\} \right) \subset \mathbb{R}^3.$$

We now find $R(\mathsf{A})$, the range space of A. This is the set of elements $\mathsf{y} \in \mathbb{R}^2$ that can be expressed as $\mathsf{y} = \mathsf{A}\mathsf{x}$ for any $\mathsf{x} \in \mathbb{R}^3$. In our case this means

$$\mathsf{y} = \mathsf{A}\mathsf{x} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 4 \end{bmatrix} x_2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} x_3.$$

What this last equation says, is that the elements of $R(\mathsf{A})$ can be expressed as a linear combination of the column vectors of matrix A. Therefore, the set $R(\mathsf{A})$ is indeed the set of all possible linear combinations of the column vectors of matrix A. We then conclude that

$$R(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\} \right).$$

Notice that

$$2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \Rightarrow \quad \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\} \right) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\} \right).$$

Therefore, we conclude that the smallest set whose span is $R(\mathsf{A})$ contains two elements, the first and third columns of A, that is,

$$R(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\} \right) = \mathbb{R}^2.$$

◁

In the Example 2.5.1 above we have seen that both sets $N(\mathsf{A})$ and $R(\mathsf{A})$ can be expressed as spans of appropriate vectors. It is not surprising to see that the same property holds for $N(\mathsf{A}^T)$ and $R(\mathsf{A}^T)$, as we observe in the following Example.

EXAMPLE 2.5.2: Find the $N(\mathsf{A}^T)$ and $R(\mathsf{A}^T)$ for the function $\mathsf{A}^T : \mathbb{R}^2 \to \mathbb{R}^3$, where A is the matrix in Example 2.5.1, that is,

$$\mathsf{A}^T = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 1 \end{bmatrix}.$$

SOLUTION: We start finding the $N(\mathsf{A}^T)$, that is, all vectors $\mathsf{y} \in \mathbb{R}^2$ solutions of the homogeneous equation $\mathsf{A}^T\mathsf{y} = \mathsf{0}$. Using the Gauss method we obtain,

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & -5 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad N(\mathsf{A}^T) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \subset \mathbb{R}^2.$$

We now find the $R(\mathsf{A}^T)$. This is the set of $\mathsf{x} \in \mathbb{R}^3$ that can be expressed as $\mathsf{x} = \mathsf{A}^T\mathsf{y}$ for any $\mathsf{y} \in \mathbb{R}^2$, that is,

$$\mathsf{x} = \mathsf{A}^T\mathsf{y} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} y_1 + \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} y_2.$$

As in the previous example, this last equation says that the elements of $R(\mathsf{A}^T)$ can be expressed as a linear combination of the column vectors of matrix A^T. Therefore, the set

$R(\mathsf{A}^T)$ is indeed the set of all possible linear combinations of the column vectors of matrix A^T. We then conclude that

$$R(\mathsf{A}^T) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \right\}\right) \subset \mathbb{R}^3.$$

In Fig. 26 we have sketched the sets $N(\mathsf{A})$ and $R(\mathsf{A}^T)$, which are subsets of $\mathbb{R}^3$. ◁

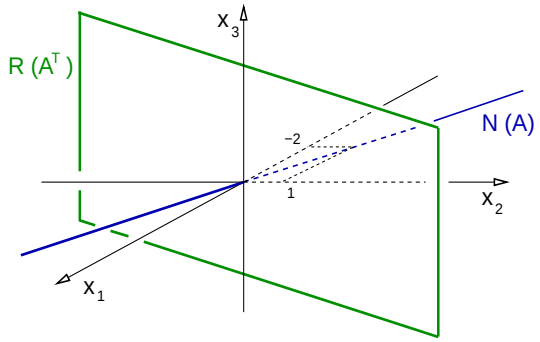

FIGURE 26. The horizontal blue line is the $N(\mathsf{A})$, and the vertical green plane is the $R(\mathsf{A}^T)$, for the matrix A is given in Examples 2.5.1 and 2.5.2. Notice that the blue line is perpendicular to the green plane. We will see that this is always true for all $m \times n$ matrices.

EXAMPLE 2.5.3: Find the $N(\mathsf{A})$ and $R(\mathsf{A})$ for the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 6 & -2 \\ 3 & 9 & -3 \end{bmatrix}.$$

SOLUTION: Any element $\mathsf{x} \in N(\mathsf{A})$ must be solution of the homogeneous equation $\mathsf{Ax} = \mathsf{0}$. The Gauss method implies,

$$\begin{bmatrix} 1 & 3 & -1 \\ 2 & 6 & -2 \\ 3 & 9 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = -3x_2 + x_3, \\ x_2, x_3 \text{ free variables.} \end{cases}$$

Therefore, every element $\mathsf{x} \in N(\mathsf{A})$ has the form

$$\boldsymbol{x} = \begin{bmatrix} -3x_2 + x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} x_3 \quad \Rightarrow \quad N(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}\right).$$

The $R(\mathsf{A})$ is the set of all $\mathsf{y} \in \mathbb{R}^3$ such that $\mathsf{y} = \mathsf{Ax}$ for some $\mathsf{x} \in \mathbb{R}^3$. Therefore,

$$\mathsf{y} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 6 & -2 \\ 3 & 9 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix},$$

and we conclude that

$$R(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} \right\}\right).$$

However, the expression above can be simplified noticing that

$$\begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} = - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

So the smallest sets whose span is $R(\mathsf{A})$ contain only one vector, and a possible choice is the following:

$$R(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\} \right).$$

◁

2.5.2. **Main properties.** The property of the null and range sets we have found for the matrix in Examples 2.5.1 and 2.5.2 also holds for every matrix. This will be an important property later on, so we give it a name.

Definition 2.5.2. *A subset of $U \subset \mathbb{F}^n$ is called **closed under linear combination** iff for all elements x, $\mathsf{y} \in U$ and all scalars a, $b \in \mathbb{F}$ holds that $(a\mathsf{x} + b\mathsf{y}) \in U$.*

In words, a set $U \subset \mathbb{F}^n$ is closed under linear combination iff every linear combination of elements in U stays in U. In Chapter 4 we generalize the linear combination structure of $\mathbb{F}^n$ into a structure we call a vector space; we will see in that Chapter that sets in a vector space which are closed under linear combinations are smaller vector spaces inside the original vector space, and will be called subspaces. We now state as a general result the property we found both in Examples 2.5.1 and 2.5.2.

Theorem 2.5.3. *The sets $N(\mathsf{A}) \subset \mathbb{F}^n$ and $R(\mathsf{A}) \subset \mathbb{F}^m$ of a matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ are closed under linear combinations in $\mathbb{F}^n$ and $\mathbb{F}^m$, respectively. Furthermore, denoting the matrix $\mathsf{A} = \big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big]$, we conclude that $R(\mathsf{A}) = \text{Span}\big(\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\}\big)$.*

It is common in the literature to introduce the *column space* of an $m \times n$ matrix $\mathsf{A} = \big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big]$, denoted as $\text{Col}(\mathsf{A})$, as the set of all linear combinations of the column vectors of A, that is, $\text{Col}(\mathsf{A}) = \text{Span}\big(\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\}\big)$. The Proposition above then says that

$$R(\mathsf{A}) = \text{Col}(\mathsf{A}).$$

Proof of Theorem 2.5.3: The sets $N(\mathsf{A})$ and $R(\mathsf{A})$ are closed under linear combinations because the matrix-vector product is a linear operation. Consider two arbitrary elements x_1, $\mathsf{x}_2 \in N(\mathsf{A})$, that is, $\mathsf{A}\mathsf{x}_1 = \mathsf{0}$ and $\mathsf{A}\mathsf{x}_2 = \mathsf{0}$. Then, for any a, $b \in \mathbb{F}$ holds

$$\mathsf{A}(a\mathsf{x}_1 + b\mathsf{x}_2) = a\,\mathsf{A}\mathsf{x}_1 + b\,\mathsf{A}\mathsf{x}_2 = \mathsf{0} \quad \Rightarrow \quad (a\mathsf{x}_1 + b\mathsf{x}_2) \in N(\mathsf{A}).$$

Therefore, $N(\mathsf{A}) \subset \mathbb{F}^n$ is closed under linear combinations. Analogously, consider two arbitrary elements y_1, $\mathsf{y}_2 \in R(\mathsf{A})$, that is, there exist x_1, $\mathsf{x}_2 \in \mathbb{F}^n$ such that $\mathsf{y}_1 = \mathsf{A}\mathsf{x}_1$ and $\mathsf{y}_2 = \mathsf{A}\mathsf{x}_2$. Then, for any a, $b \in \mathbb{F}$ holds

$$(a\mathsf{y}_1 + b\mathsf{y}_2) = a\,\mathsf{A}\mathsf{x}_1 + b\,\mathsf{A}\mathsf{x}_2 = \mathsf{A}(a\mathsf{x}_1 + b\mathsf{x}_2) \quad \Rightarrow \quad (a\mathsf{y}_2 + b\mathsf{y}_2) \in R(\mathsf{A}).$$

Therefore, $R(\mathsf{A}) \subset \mathbb{F}^m$ is closed under linear combinations. The furthermore part is proved as follows. Denote $\mathsf{A} = \big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big]$, then any element $\mathsf{y} \in R(\mathsf{A})$ can be expressed as $\mathsf{y} = \mathsf{A}\mathsf{x}$ for some $\mathsf{x} = [x_i] \in \mathbb{F}^n$, that is,

$$\mathsf{y} = \mathsf{A}\mathsf{x} = \big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathsf{A}_{:1}\, x_1 + \cdots + \mathsf{A}_{:n}\, x_n \in \text{Span}\big(\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\}\big).$$

This implies that $R(\mathsf{A}) \subset \text{Col}(\mathsf{A})$. The opposite inclusion, $\text{Col}(\mathsf{A}) \subset R(\mathsf{A})$ is trivial, since any element of the form $\mathsf{y} = \mathsf{A}_{:1}\, x_1 + \cdots + \mathsf{A}_{:n}\, x_n \in \text{Col}(\mathsf{A})$ also belongs to $R(\mathsf{A})$, since $\mathsf{y} = \mathsf{A}\mathsf{x}$, where $\mathsf{x} = [x_i]$. This establishes the Theorem. □

The null and range spaces of a square matrix characterize whether the matrix is invertible or not. The following result is a simple rewriting of Theorem 2.4.3.

Theorem 2.5.4. *Given a matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, the following statements are equivalent:*

(a) The matrix A^{-1} exists;

(b) $N(\mathsf{A}) = \{\mathsf{0}\}$;

(c) $R(\mathsf{A}) = \mathbb{F}^n$.

Proof of Theorem 2.5.4: This result follows from Theorem 2.4.3. More precisely, part (b) follows from Theorem 2.4.3 part (c). While part (c) follows from Theorem 2.4.3 part (e). □

2.5.3. **Gauss operations.** In this section we study the relations between the four spaces associated with the matrices A and B in the case that matrix B can be obtained from matrix A by performing Gauss operations. We use the following notation: $\mathsf{A} \overset{\text{row}}{\longleftrightarrow} \mathsf{B}$ to indicate that A can be transformed into matrix B by performing Gauss operations on the rows of A.

Theorem 2.5.5. *If matrices A, $\mathsf{B} \in \mathbb{F}^{m,n}$, then*

(a) $\mathsf{A} \overset{row}{\longleftrightarrow} \mathsf{B} \quad \Leftrightarrow \quad N(\mathsf{A}) = N(\mathsf{B})$;

(b) $\mathsf{A} \overset{row}{\longleftrightarrow} \mathsf{B} \quad \Leftrightarrow \quad R(\mathsf{A}^T) = R(\mathsf{B}^T)$.

It is simple to see that this result is true. Gauss operations do not change the solutions of linear systems, so the linear system $\mathsf{A}\mathsf{x} = \mathsf{0}$ has exactly the same solutions x as the linear system $\mathsf{B}\mathsf{x} = \mathsf{0}$, that is, $N(\mathsf{A}) = N(\mathsf{B})$. The second property must be also true, since Gauss operations on the rows of A are equivalent to Gauss operations on the columns of A^T. Now, it is simple to see that each of the Gauss operations on the columns of A^T do not change the $\text{Col}(\mathsf{A}^T)$, hence $R(\mathsf{A}^T) = R(\mathsf{B}^T)$.

One could say that the paragraph above is enough for a proof of the Theorem. However, we would like to have a more detailed presentation of the ideas above. One way is to use matrix multiplication to express the Gauss operation property that they do not change the solutions to linear systems. One can prove the following: If the $m \times n$ matrices A and B are related by Gauss operations on their rows, then there exists an $m \times m$ invertible matrix G such that $\mathsf{G}\mathsf{A} = \mathsf{B}$. The proof of this property is simple. Each one of the Gauss operations is associated with an invertible matrix, E, called an elementary Gauss matrix. Every elementary Gauss matrix is invertible, since every Gauss operation can always be reversed. The result of several Gauss operations on a matrix A is the product of the appropriate elementary Gauss matrices in the same order as the Gauss operations are performed. If the matrix B is obtained from matrix A by doing Gauss operations given by matrices E_i, for $i = 1, \cdots, k$, in that order, we can express the result of the Gauss method as follows:

$$\mathsf{E}_k \cdots \mathsf{E}_1 \mathsf{A} = \mathsf{B}, \quad \mathsf{G} = \mathsf{E}_k \cdots \mathsf{E}_1 \quad \Rightarrow \quad \mathsf{G}\mathsf{A} = \mathsf{B}.$$

Since each elementary Gauss matrix is invertible, then the matrix G is also invertible. The following example shows all 3×3 elementary Gauss matrices.

EXAMPLE 2.5.4: Find all the elementary Gauss matrices which operate on $3 \times n$ matrices.

SOLUTION: In the case of $3 \times n$ matrices, the elementary Gauss matrices are 3×3. We present these matrices for each one of the three main types of Gauss operations. Consider the matrices E_i for $i = 1, 2, 3$ are given by

$$\mathsf{E}_1 = \begin{bmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & k \end{bmatrix};$$

then, the product $\mathsf{E}_i A$ represents the Gauss operation of multiplying by k the first, second and third row of A, respectively. Consider the matrices E_i for $i = 4, 5, 6$ given by

$$\mathsf{E}_4 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_5 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathsf{E}_6 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix};$$

then, the product $\mathsf{E}_i A$ for $i = 4, 5, 6$ represents the Gauss operation of interchanging the first and second, the first and third, and the second and third rows of A, respectively. Finally, consider the matrices E_i for $j = 7, \cdots, 12$ given by

$$\mathsf{E}_7 = \begin{bmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_8 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_9 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix},$$

$$\mathsf{E}_{10} = \begin{bmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_{11} = \begin{bmatrix} 1 & 0 & k \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix};$$

then, the product $\mathsf{E}_i \mathsf{A}$ for $i = 7, \cdots, 12$ represents the Gauss operation of multiplying by k one row of A and add the result to another row of A. ◁

Proof of Theorem 2.5.5: Recall the comment below Theorem 2.5.5: If the $m \times n$ matrices A and B are related by Gauss operations on their rows, then there exits an $m \times m$ invertible matrix G such that $\mathsf{GA} = \mathsf{B}$. This observation is the key to show that $N(\mathsf{A}) = N(\mathsf{B})$, since given any element $\mathsf{x} \in N(\mathsf{A})$

$$\mathsf{Ax} = \mathsf{0} \quad \Leftrightarrow \quad \mathsf{GAx} = \mathsf{0},$$

where the equivalence follows from G being invertible. Then it is simple to see that

$$\mathsf{0} = \mathsf{GAx} = \mathsf{Bx} \quad \Leftrightarrow \quad \mathsf{x} \in N(\mathsf{B}).$$

Therefore, we have shown that $N(\mathsf{A}) = N(\mathsf{B})$.

We now show the converse statement. If $N(\mathsf{A}) = N(\mathsf{B})$ this means that their reduced echelon forms are the same, that is, $\mathsf{E_A} = \mathsf{E_B}$. This means that there exist Gauss operations on the rows of A that transform it into matrix B.

We now show that $R(\mathsf{A}^T) = R(\mathsf{B}^T)$. Given any element $\mathsf{x} \in R(\mathsf{A}^T)$ we know that there exists an element $\mathsf{y} \in \mathbb{F}^m$ such that

$$\mathsf{x} = \mathsf{A}^T\mathsf{y} = \mathsf{A}^T\mathsf{G}^T\big(\mathsf{G}^T\big)^{-1}\mathsf{y} = (\mathsf{GA})^T\tilde{\mathsf{y}} = \mathsf{B}^T\tilde{\mathsf{y}}, \qquad \tilde{\mathsf{y}} = \big(\mathsf{G}^T\big)^{-1}\mathsf{y}.$$

We have shown that given any $\mathsf{x} \in R(\mathsf{A}^T)$, then $\mathsf{x} \in R(\mathsf{B}^T)$, that is, $R(\mathsf{A}^T) \subset R(\mathsf{B}^T)$. The opposite implication is proven in the same way: Given any $\mathsf{x} \in R(\mathsf{B}^T)$ there exists $\tilde{\mathsf{y}} \in \mathbb{F}^m$ such that

$$\mathsf{x} = \mathsf{B}^T\tilde{\mathsf{y}} = \mathsf{B}^T\big(\mathsf{G}^T\big)^{-1}\mathsf{G}^T\tilde{\mathsf{y}} = \big(\mathsf{G}^{-1}\mathsf{B}\big)^T\mathsf{y} = \mathsf{A}^T\mathsf{y}, \qquad \mathsf{y} = \mathsf{G}^T\tilde{\mathsf{y}}.$$

We have shown that given any $\mathsf{x} \in R(\mathsf{B}^T)$, then $\mathsf{x} \in R(\mathsf{A}^T)$, that is, $R(\mathsf{B}^T) \subset R(\mathsf{A}^T)$. Therefore, $R(\mathsf{A}^T) = R(\mathsf{B}^T)$.

We now show that the converse statement. Assume that $R(\mathsf{A}^T) = R(\mathsf{B}^T)$. This means that every row in matrix A is a linear combination of the rows in matrix B. This also means that there exists Gauss operations on the rows of A such that transform A into B. This establishes the Theorem. □

EXAMPLE 2.5.5: Verify Theorem 2.5.5 for matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ and $\mathsf{E_A} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

SOLUTION: Since $\mathsf{E_A}$ is the reduced echelon form of matrix A, then $R(\mathsf{A}^T) = R\big((\mathsf{E_A})^T\big)$. Consider the matrix A and its reduced echelon form $\mathsf{E_A}$ given by

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}, \qquad \mathsf{E_A} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Their respective transposed matrices are given by

$$\mathsf{A}^T = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 1 \end{bmatrix}, \qquad (\mathsf{E_A})^T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

The result of Theorem 2.5.5 is that

$$R(\mathsf{A}^T) = R\big((\mathsf{E_A})^T\big) \quad \Leftrightarrow \quad \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \right\}\right) = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}\right).$$

We can verify that this result above is correct, since the column vectors in A^T are linear combinations of the column vectors in $(\mathsf{E_A})^T$, as the following equations show,

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

◁

The Example 2.5.5 above helps understand one important meaning of Gauss operations. *Gauss operations on the matrix A rows are linear combination of the matrix A^T columns.* This is the reason why the interchange change $\mathsf{A} \leftrightarrow \mathsf{B}$ by doing Gauss operations on rows leaves the the range spaces of their transposes invariant, that is, $R(\mathsf{A}^T) = R(\mathsf{B}^T)$.

EXAMPLE 2.5.6: Given the matrices $\mathsf{A} = \begin{bmatrix} 1 & 1 & 5 \\ 2 & 0 & 6 \\ 1 & 2 & 7 \end{bmatrix}$, $\mathsf{B} = \begin{bmatrix} 1 & -4 & 4 \\ 4 & -8 & 6 \\ 0 & -4 & 5 \end{bmatrix}$, verify whether

$$R(\mathsf{A}^T) = R(\mathsf{B}^T)? \quad R(\mathsf{A}) = R(\mathsf{B})? \quad N(\mathsf{A}) = N(\mathsf{B})? \quad N(\mathsf{A}^T) = N(\mathsf{B}^T)?$$

SOLUTION: We base our answer in Theorem 2.5.5 and an extra observation. First, let $\mathsf{E_A}$ and $\mathsf{E_B}$ be the reduced echelon forms of A and B, respectively, and let $\mathsf{E}_{\mathsf{A}^T}$ and $\mathsf{E}_{\mathsf{B}^T}$ be the reduced echelon forms of A^T and B^T respectively. The extra observation is the following: $\mathsf{E_A} = \mathsf{E_B}$ iff $\mathsf{A} \overset{\text{row}}{\longleftrightarrow} \mathsf{B}$. This observation and Theorem 2.5.5 imply that $\mathsf{E_A} = \mathsf{E_B}$ is equivalent to $R(\mathsf{A}^T) = R(\mathsf{B}^T)$ and it is also equivalent to $N(\mathsf{A}) = N(\mathsf{B})$. We then find $\mathsf{E_A}$ and $\mathsf{E_B}$,

$$\mathsf{A} = \begin{bmatrix} 1 & 1 & 5 \\ 2 & 0 & 6 \\ 1 & 2 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 5 \\ 0 & -2 & -4 \\ 0 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \mathsf{E_A},$$

$$\mathsf{B} = \begin{bmatrix} 1 & -4 & 4 \\ 4 & -8 & 6 \\ 0 & -4 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & -4 & 4 \\ 0 & 8 & -10 \\ 0 & -4 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & -4 & 4 \\ 0 & 4 & -5 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -5/4 \\ 0 & 0 & 0 \end{bmatrix} = \mathsf{E_B}.$$

Since $\mathsf{E_A} \neq \mathsf{E_B}$, we conclude that $R(\mathsf{A}^T) \neq R(\mathsf{B}^T)$. This result also says that $N(\mathsf{A}) \neq N(\mathsf{B})$. So for the first and third questions, the answer is no.

A similar argument also says that $\mathsf{E}_{\mathsf{A}^T} = \mathsf{E}_{\mathsf{B}^T}$ iff $\mathsf{A}^T \overset{\text{row}}{\longleftrightarrow} \mathsf{B}^T$. This observation and Theorem 2.5.5 imply that $\mathsf{E}_{\mathsf{A}^T} = \mathsf{E}_{\mathsf{B}^T}$ is equivalent to $R(\mathsf{A}) = R(\mathsf{B})$ and it is also equivalent

to $N(A^T) = N(B^T)$. We then find E_{A^T} and E_{B^T},

$$A^T = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 2 \\ 5 & 6 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 1 \\ 0 & -4 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \end{bmatrix} = E_{A^T},$$

$$B^T = \begin{bmatrix} 1 & 4 & 0 \\ -4 & -8 & -4 \\ 4 & 6 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 4 & 0 \\ 0 & 8 & -4 \\ 0 & -10 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 4 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \end{bmatrix} = E_{B^T}.$$

Since $E_{A^T} = E_{B^T}$, we conclude that $R(A) = R(B)$. This result on the reduced echelon forms also says that $N(A^T) = N(B^T)$. So for the second and four questions, the answer is yes. ◁

We finish this Section with two results concerning the ranks of a matrix and its transpose. The first result says that transposition operation on a matrix does not change its rank.

Theorem 2.5.6. *Every matrix $A \in \mathbb{F}^{m,n}$ satisfies that* $\text{rank}(A) = \text{rank}(A^T)$.

We delay the proof of the first result to Chapter 6, where we introduce the notion of inner product in a vector space. Using an inner product in $\mathbb{F}^n$, the proof of Theorem 2.5.6 below is simple. We now introduce a particular name for those matrices having the maximum possible rank.

Definition 2.5.7. *An $m \times n$ matrix A has* ***full rank*** *iff* $\text{rank}(A) = \min(m, n)$.

Our last result of this Section concerns full rank matrices and relates the rank of a matrix A to the size of both $N(A)$ and $N(A^T)$.

Theorem 2.5.8. *If matrix $A \in \mathbb{F}^{m,n}$ has full rank, then hold:*

(a) If $m = n$, then $\text{rank}(A) = \text{rank}(A^T) = n = m \quad \Leftrightarrow \quad \{0\} = N(A) = N(A^T) \subset \mathbb{F}^n$;

(b) If $m < n$, then $\text{rank}(A) = \text{rank}(A^T) = m < n \quad \Leftrightarrow \quad \begin{cases} \{0\} \subsetneq N(A) \subset \mathbb{F}^n, \\ \{0\} = N(A^T) \subset \mathbb{F}^m; \end{cases}$

(c) If $m > n$, then $\text{rank}(A) = \text{rank}(A^T) = n < m \quad \Leftrightarrow \quad \begin{cases} \{0\} = N(A) \subset \mathbb{F}^n, \\ \{0\} \subsetneq N(A^T) \subset \mathbb{F}^m. \end{cases}$

Proof of Theorem 2.5.8: We start recaling that the rank of a matrix A is the number of pivot columns in its reduced echelon form E_A.

If an $m \times n$ matrix A has $\text{rank}(A) = n$, this means two things: First, $n \leqslant m$; and second, that every column in E_A has a pivot. The latter property implies that there is no free variables in the solution of the equation $Ax = 0$, and so $x = 0$ is the unique solution. We conclude that $N(A) = \{0\}$. In order to study $N(A^T)$ we need to consider the two possible cases $n = m$ or $n < m$. If $n = m$, then the matrices A and A^T are square, and the same argument about free variables applies to solutions of $A^T y = 0$, so we conclude that $N(A^T) = \{0\}$. This proves (a). In the case that $n < m$, then there are free variables in the solution of the equation $A^T y = 0$, therefore $\{0\} \subsetneq N(A^T)$. This proves (c).

If an $m \times n$ matrix A has $\text{rank}(A) = m$, recalling that $\text{rank}(A) = \text{rank}(A^T)$, this means two things: First, $m \leqslant n$; and second, that every column in E_{A^T} has a pivot. The latter property shows that A^T is full rank, so the argument above shows that $N(A^T) = \{0\}$. Since the case $n = m$ has already been studied, so we only need to consider the case $m < n$. In this case there are free variables in the solution to the equation $Ax = 0$, therefore $\{0\} \subsetneq N(A)$. This proves (b), and we have established the Theorem. □

Further reading. Section 4.2 in Meyer's book [3] follows closely the descriptions of the four spaces given here, although in more depth.

2.5.4. **Exercises.**

2.5.1.- Find $R(\mathsf{A})$ and $R(\mathsf{A}^T)$ for the matrix A below and express them as the span of the smallest possible set of vectors, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{bmatrix}.$$

2.5.2.- Find the $N(\mathsf{A})$, $R(\mathsf{A})$, $N(\mathsf{A}^T)$ and $R(\mathsf{A}^T)$ for the matrix A below and express them as the span of the smallest possible set of vectors, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 & 1 & 5 \\ -2 & -4 & 0 & 4 & -2 \\ 1 & 2 & 2 & 4 & 9 \end{bmatrix}.$$

2.5.3.- Let A be a 3×3 matrix such that

$$R(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \Big\}\Big),$$

$$N(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \Big\}\Big).$$

(a) Show that the linear system $\mathsf{Ax} = \mathsf{b}$ is consistent for

$$\mathsf{b} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}.$$

(b) Show that the system $\mathsf{Ax} = \mathsf{b}$ above has infinitely many solutions.

2.5.4.- Prove:

(a) $\mathsf{Ax} = \mathsf{b}$ is consistent iff $\mathsf{b} \in R(\mathsf{A})$.

(b) The consistent system $\mathsf{Ax} = \mathsf{b}$ has a unique solution iff $N(\mathsf{A}) = \{\mathsf{0}\}$.

2.5.5.- Prove: A matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ is invertible iff $R(\mathsf{A}) = \mathbb{F}^n$.

2.5.6.- Let $\mathsf{A} \in \mathbb{F}^{n,n}$ be an invertible matrix. Find the spaces $N(\mathsf{A})$, $R(\mathsf{A})$, $N(\mathsf{A}^T)$, and $R(\mathsf{A}^T)$.

2.5.7.- Consider the matrices

$$\mathsf{A} = \begin{bmatrix} 1 & 5 & 3 \\ 2 & 1 & -3 \\ 1 & 3 & 1 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Answer the questions below. If the answer is *yes*, give a proof; if the answer is *no*, give a counter-example.

(a) Is $R(\mathsf{A}) = R(\mathsf{B})$?

(b) Is $R(\mathsf{A}^T) = R(\mathsf{B}^T)$?

(c) Is $N(\mathsf{A}) = N(\mathsf{B})$?

(d) Is $N(\mathsf{A}^T) = N(\mathsf{B}^T)$?

2.6. LU-factorization

A factorization of a number means to decompose that number as a product of appropriate factors. For example, the prime factorization of an integer number means to decompose that integer as a product of prime numbers, like $70 = (2)(5)(7)$. In a similar way, a factorization of a matrix means to decompose that matrix, using matrix multiplication, as a product of appropriate factors. In this Section we introduce a particular type of factorization, called LU-factorization, which stands for *lower triangular-upper triangular* factorization. A given matrix A is expressed as a product $\mathsf{A} = \mathsf{LU}$, where L is lower triangular and U is upper triangular. This type of factorization can be useful to solve systems of linear equations having A as the coefficient matrix. The LU-factorization of A reduces the number of algebraic operations needed to solve the linear system, saving computer time when solving these systems using a computer. However, not every matrix has this type of factorization. We provide sufficient conditions on A that guarantee its LU-factorization.

2.6.1. Main definitions.

We start with few basic definitions.

Definition 2.6.1. *An $m \times n$ matrix is called **upper triangular** iff every element below the diagonal vanishes, and **lower triangular** iff every element above the diagonal vanishes.*

Example 2.6.1: The matrices U_1 and U_2 below are upper triangular, while L_1 and L_2 are lower triangular,

$$\mathsf{U}_1 = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{U}_2 = \begin{bmatrix} 2 & 3 & 4 & 3 \\ 0 & 5 & 6 & 2 \\ 0 & 0 & 1 & 5 \end{bmatrix}, \quad \mathsf{L}_1 = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 4 & 0 \\ 5 & 6 & 7 \end{bmatrix}, \quad \mathsf{L}_2 = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 5 & 6 & 7 & 0 \end{bmatrix}.$$

◁

Definition 2.6.2. *An $m \times n$ matrix A has an **LU-factorization** iff there exists a square $m \times m$ lower triangular matrix L and an $m \times n$ upper triangular matrix U such that $\mathsf{A} = \mathsf{LU}$.*

In the particular case that A is a square matrix both L and U are square matrices.

Example 2.6.2: Verify that matrix $\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ and matrix $\mathsf{U} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ are the LU-factorization of matrix $\mathsf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$.

Solution: We need to verify that $\mathsf{A} = \mathsf{LU}$. This is indeed the case, since

$$\mathsf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \mathsf{A},$$

that is, $\mathsf{A} = \mathsf{LU}$. ◁

Example 2.6.3: Verify that matrix $\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -3 & 1 \end{bmatrix}$ and matrix $\mathsf{U} = \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ are the LU-factorization of matrix $\mathsf{A} = \begin{bmatrix} 2 & 4 & -1 \\ -4 & -5 & 3 \\ 2 & -5 & -4 \end{bmatrix}$.

SOLUTION: We need to verify that $\mathsf{A} = \mathsf{LU}$. This is the case, since

$$\mathsf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -1 \\ -4 & -5 & 3 \\ 2 & -5 & -4 \end{bmatrix} = \mathsf{A} \quad \Rightarrow \quad \mathsf{A} = \mathsf{LU},$$

◁

EXAMPLE 2.6.4: Verify that matrix $\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}$ and matrix $\mathsf{U} = \begin{bmatrix} 1 & 4 \\ 0 & -3 \\ 0 & 0 \end{bmatrix}$ are the LU-factorization of matrix $\mathsf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.

SOLUTION: We need to verify that $\mathsf{A} = \mathsf{LU}$. This is the case, since

$$\mathsf{LU} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & -3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \mathsf{A} \quad \Rightarrow \quad \mathsf{A} = \mathsf{LU}.$$

◁

EXAMPLE 2.6.5: Verify that matrix $\mathsf{L} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ and matrix $\mathsf{U} = \begin{bmatrix} 1 & 3 & 5 \\ 0 & -2 & -4 \end{bmatrix}$ are the LU-factorization of matrix $\mathsf{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$.

SOLUTION: We need to verify that $\mathsf{A} = \mathsf{LU}$. This is the case, since

$$\mathsf{LU} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 0 & -2 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} = \mathsf{A} \quad \Rightarrow \quad \mathsf{A} = \mathsf{LU}.$$

◁

2.6.2. **A sufficient condition.** We now provide sufficient conditions on a matrix that imply that such matrix has an LU-factorization.

Theorem 2.6.3. *If the $m \times n$ matrix A can be transformed into an echelon form without row exchanges and without row rescaling, then A has an LU-factorization, $\mathsf{A} = \mathsf{LU}$. The matrix U is the $m \times n$ echelon form mentioned above. Matrix $\mathsf{L} = [L_{ij}]$ is an $m \times m$ lower triangular matrix satisfying two conditions: First, the coefficients $L_{ii} = 1$; second, the coefficients L_{ij} below the diagonal are equal to the multiple of row j that is subtracted from row i in other to annihilate the (i, j) position during the Gauss elimination method.*

This result is usually found in the literature for the case of square matrices, $m = n$. A reason for this is that more often than not only square matrices appear in applications, like finding solutions of a $n \times n$ linear system $\mathsf{Ax} = \mathsf{b}$, where matrix A is not only square but also invertible. We have decided in these notes to present the most general version of the LU-factorization in the statement above without proof, and we sketch a proof in the case of 2×2 and 3×3 matrices only. Before this sketch of a proof we show few examples in order to have a better understanding of the statement in Theorem 2.6.3.

EXAMPLE 2.6.6: Find the LU-factorization of matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

SOLUTION: Theorem 2.6.3 says that U is an echelon form of A obtained without row exchanges and without row rescaling, while L has the form $\mathsf{L} = \begin{bmatrix} 1 & 0 \\ L_{21} & 1 \end{bmatrix}$, that is, we need to

find only the coefficient L_{21}. In this simple example we can summarize the whole procedure in the following diagram:

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{\text{row}(2)-\mathbf{3}\,\text{row}(1)} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} = U, \qquad \mathsf{L} = \begin{bmatrix} 1 & 0 \\ \mathbf{3} & 1 \end{bmatrix}.$$

This is what we have done: An echelon form for A is obtained with only one Gauss operation: Subtract from the second row the first row multiplied by 3. The resulting echelon form of A is matrix U already. And $L_{21} = 3$, since we multiplied by 3 the first row and we subtracted it from the second row. So we have both U and L, and the result is:

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} = \mathsf{LU}.$$

◁

EXAMPLE 2.6.7: Find the LU-factorization of the matrix $\mathsf{A} = \begin{bmatrix} 2 & 4 & -1 \\ -4 & -5 & 3 \\ 2 & -5 & -4 \end{bmatrix}$.

SOLUTION: We recall that we have to find the coefficients below the diagonal in matrix

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}.$$

From the first Gauss operation we obtain L_{21} as follows:

$$\begin{bmatrix} 2 & 4 & -1 \\ -4 & -5 & 3 \\ 2 & -5 & -4 \end{bmatrix} \xrightarrow{\text{row}(2)-(-\mathbf{2})\,\text{row}(1)} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 2 & -5 & -4 \end{bmatrix} \quad \Rightarrow \quad \mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-2} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}.$$

From the second Gauss operation we obtain L_{31} as follows:

$$\begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 2 & -5 & -4 \end{bmatrix} \xrightarrow{\text{row}(3)-\mathbf{1}\,\text{row}(1)} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & -9 & -3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ \mathbf{1} & L_{32} & 1 \end{bmatrix}.$$

From the third Gauss operation we obtain L_{32} as follows:

$$\begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & -9 & -3 \end{bmatrix} \xrightarrow{\text{row}(3)-(-\mathbf{3})\,\text{row}(2)} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & \mathbf{-3} & 1 \end{bmatrix}.$$

We then conclude,

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -3 & 1 \end{bmatrix}, \qquad \mathsf{U} = \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and we have the decomposition

$$\mathsf{A} = \begin{bmatrix} 2 & 4 & -1 \\ -4 & -5 & 3 \\ 2 & -5 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \mathsf{LU}.$$

◁

In the following Example we show a matrix A that does not have an LU-factorization. The reason is that in the procedure to find U appears a diagonal coefficient that vanishes. Since row interchanges are prohibited, then there is no LU-factorization in this case.

EXAMPLE 2.6.8: Show that matrix $\mathsf{A} = \begin{bmatrix} 2 & 4 & -1 \\ -4 & -8 & 3 \\ 2 & -5 & -4 \end{bmatrix}$ has no LU-factorization.

SOLUTION: From the first Gauss operation we obtain L_{21} as follows:

$$\begin{bmatrix} 2 & 4 & -1 \\ -4 & -8 & 3 \\ 2 & -5 & -4 \end{bmatrix} \xrightarrow{\text{row}(2)-(-\mathbf{2})\,\text{row}(1)} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 0 & 1 \\ 2 & -5 & -4 \end{bmatrix} \quad \Rightarrow \quad \mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -\mathbf{2} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}.$$

From the second Gauss operation we obtain L_{31} as follows:

$$\begin{bmatrix} 2 & 4 & -1 \\ 0 & 3 & 1 \\ 2 & -5 & -4 \end{bmatrix} \xrightarrow{\text{row}(3)-\mathbf{1}\,\text{row}(1)} \begin{bmatrix} 2 & 4 & -1 \\ 0 & 0 & 1 \\ 0 & -9 & -3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ \mathbf{1} & L_{32} & 1 \end{bmatrix}.$$

However, we cannot continue the Gauss method to find an echelon form for A without interchanging the second and third rows. Therefore, matrix A has no LU-factorization. ◁

In the Example 2.6.7 we also had a diagonal coefficient in matrix U that vanishes, the coefficient $U_{33} = 0$. However, in this case A did have an LU-factorization, since this vanishing coefficient was in the last row of matrix U, and no further Gauss operations were needed.

Proof of Theorem 2.6.3: We only give a proof in the case that matrix A is 2×2 or 3×3. This would give an idea how to construct a proof in the general case. This generalization does not involve new ideas, only a more sophisticated notation.

Assume that matrix A is 2×2, that is

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Matrix A is assumed to satisfy the following property: A can be transformed into echelon form by the only Gauss operation of multiplying a row and add that result to another row. If the coefficient $A_{21} = 0$, then matrix A is upper triangular already and it has the trivial LU-factorization $\mathsf{A} = \mathsf{I}_2\,\mathsf{A}$. If the coefficient $A_{21} \neq 0$, then from Example 2.6.8 we know that the assumption on A implies that the coefficient $A_{11} \neq 0$. Then, we can perform the Gauss operation $\mathsf{EA} = \mathsf{U}$, that is

$$\mathsf{EA} = \begin{bmatrix} 1 & 0 \\ -\dfrac{A_{21}}{A_{11}} & 1 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & \dfrac{A_{22}A_{11} - A_{21}A_{12}}{A_{11}} \end{bmatrix} = \mathsf{U}.$$

Matrix U is the upper triangular, and denoting

$$U_{22} = \frac{A_{22}A_{11} - A_{21}A_{12}}{A_{11}}, \qquad L_{21} = \frac{A_{21}}{A_{11}},$$

we obtain that

$$\mathsf{U} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & U_{22} \end{bmatrix}, \quad \mathsf{E} = \begin{bmatrix} 1 & 0 \\ -L_{21} & 1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{E}^{-1} = \begin{bmatrix} 1 & 0 \\ L_{21} & 1 \end{bmatrix}.$$

We conclude that matrix A has an LU-factorization

$$\mathsf{A} = \begin{bmatrix} 1 & 0 \\ L_{21} & 1 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ 0 & U_{22} \end{bmatrix}.$$

Assume now that matrix A is 3×3, that is

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}.$$

Once again, matrix A is assumed to satisfy the following property: A can be transformed into echelon form by the only Gauss operation of multiplying a row and add that result to another row. If any of the coefficients A_{21} or A_{13} is non-zero, then the assumption on A implies that the coefficient $A_{11} \neq 0$. Then, we can perform the Gauss operation $\mathsf{E}_2\mathsf{E}_1\mathsf{A} = \mathsf{B}$, that is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -L_{31} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -L_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{bmatrix},$$

where

$$\mathsf{E}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -L_{31} & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -L_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we used the notation

$$L_{21} = \frac{A_{21}}{A_{11}}, \quad L_{31} = \frac{A_{31}}{A_{11}},$$

and the B_{ij} coefficients can be easily computed. Now, if the coefficient $B_{32} \neq 0$, then the assumption on A implies that the coefficient $B_{22} \neq 0$. We then assume that $B_{22} \neq 0$ and we proceed one more step. We can perform the Gauss operation $\mathsf{E}_3\,\mathsf{B} = \mathsf{U}$, that is

$$\mathsf{E}_3\,\mathsf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -L_{31} & 1 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & B_{22} & B_{23} \\ 0 & 0 & U_{33} \end{bmatrix} = \mathsf{U},$$

where we used the notation $L_{31} = \dfrac{B_{32}}{B_{22}}$ and the coefficient U_{33} can be easily computed. This product can be expressed as follows,

$$\mathsf{E}_3\,\mathsf{E}_2\,\mathsf{E}_1\,\mathsf{A} = \mathsf{U} \quad \Rightarrow \quad \mathsf{A} = \mathsf{E}_1^{-1}\,\mathsf{E}_2^{-1}\,\mathsf{E}_3^{-1}\,\mathsf{U}.$$

It is not difficult to see that

$$\mathsf{E}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ L_{31} & 0 & 1 \end{bmatrix}, \quad \mathsf{E}_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & L_{32} & 1 \end{bmatrix},$$

which implies that

$$\mathsf{E}_1^{-1}\,\mathsf{E}_2^{-1}\,\mathsf{E}_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix} = \mathsf{L}.$$

We have then shown that matrix A has the LU-factorization $\mathsf{A} = \mathsf{L}\,\mathsf{U}$, where

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}, \qquad \mathsf{U} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & B_{22} & B_{23} \\ 0 & 0 & U_{33} \end{bmatrix}.$$

It is not difficult to see that in the case where the coefficient $B_{32} = 0$, then the expression $\mathsf{A} = \left(\mathsf{E}_1^{-1}\,\mathsf{E}_2^{-1}\right)\mathsf{B}$ is already the LU-factorization of matrix A. This establishes the Theorem for 2×2 and 3×3 matrices. □

2.6.3. **Solving linear systems.** Suppose we want to solve a linear system with coefficient matrix A, and suppose that this matrix has an LU-factorization. This extra information is useful to save computational time when solving the linear system. We state how this can be done in the following result.

Theorem 2.6.4. *Fix an $m \times n$ matrix A having an LU-factorization $\mathsf{A} = \mathsf{LU}$ and a vector $\mathsf{b} \in \mathbb{F}^m$. The vector $\mathsf{x} \in \mathbb{F}^n$ is solution of the linear system $\mathsf{Ax} = \mathsf{b}$ iff holds*

$$\mathsf{Ux} = \mathsf{y}, \quad \text{where} \quad \mathsf{Ly} = \mathsf{b}. \tag{2.8}$$

First solve the second equation for y, and then use this vector y to solve for vector x. This is faster than solving directly for x in the original equation $\mathsf{Ax} = \mathsf{b}$. The reason is that forward substitution can be used to solve for y, since L is lower triangular. Then, backward substitution can be used to solve for x, since U is upper triangular.

Proof of Theorem 2.6.4: Replace on the second equation in (2.8) the vector y defined by the first equation in (2.8). Hence, the vector x is solution of the system in (2.8) iff holds

$$\mathsf{L}(\mathsf{Ux}) = \mathsf{b}.$$

Since $\mathsf{A} = \mathsf{LU}$, the equation above is equivalent to $\mathsf{Ax} = \mathsf{b}$. This establishes the Theorem. □

EXAMPLE 2.6.9: Use the LU-factorization of matrix A below, to find the solution x to the system $\mathsf{Ax} = \mathsf{b}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

SOLUTION: In Example 2.6.2 we have shown that $\mathsf{A} = \mathsf{LU}$ with

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \qquad \mathsf{U} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

We first find y solution of the system $\mathsf{Ly} = \mathsf{b}$,then, having y, we find x solution of $\mathsf{Ux} = \mathsf{y}$. For the first system we have

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 3 \end{array}\right] \quad \Rightarrow \quad \left\{\begin{array}{r} y_1 = 1, \\ y_1 + y_2 = 2, \\ y_1 + y_2 + y_3 = 3, \end{array}\right\} \quad \Rightarrow \quad \mathsf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

This system is solved for y using forward substitution. For the second system we have

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right] \quad \Rightarrow \quad \left\{\begin{array}{r} x_1 + x_2 + x_3 = 1, \\ x_2 + x_3 = 1, \\ x_3 = 1, \end{array}\right\} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

This system is solved for x using back substitution. ◁

Further reading. There exists a vast literature on matrix factorization. See Section 3.10 in Meyer's book [3] for few types of generalizations of the LU-factorizations, for example, admitting row interchanges in the process to obtain the factorization, or the LDU-factorization.

2.6.4. **Exercises.**

2.6.1.- Find the LU-factorization of matrix

$$\mathsf{A} = \begin{bmatrix} 5 & 2 \\ -15 & -3 \end{bmatrix}.$$

2.6.2.- Find the LU-factorization of matrix

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 6 & 7 \end{bmatrix}.$$

2.6.3.- Find the LU-factorization of matrix

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 2 \\ 4 & 5 & 5 \\ 6 & -3 & 5 \end{bmatrix}.$$

2.6.4.- Determine if the matrix below has an LU-factorization,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 4 & 17 \\ 3 & 6 & -12 & 3 \\ 2 & 3 & -3 & 2 \\ 0 & 2 & -2 & 6 \end{bmatrix}.$$

2.6.5.- Find the LU-factorization of a tridiagonal matrix

$$\mathsf{T} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

2.6.6.- Use the LU-factorization to find the solutions to the system $\mathsf{Ax} = \mathsf{b}$, where

$$\mathsf{A} = \begin{bmatrix} 2 & 2 & 2 \\ 4 & 7 & 7 \\ 6 & 18 & 22 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 12 \\ 24 \\ 12 \end{bmatrix}.$$

2.6.7.- Find the values of the number c such that matrix A below has no LU-factorization,

$$\mathsf{A} = \begin{bmatrix} c & 2 & 0 \\ 1 & c & 1 \\ 0 & 1 & c \end{bmatrix}.$$

CHAPTER 3. DETERMINANTS

3.1. DEFINITIONS AND PROPERTIES

A determinant is a scalar associated to a square matrix which can be used to determine whether the matrix is invertible or not, and this property is the origin of its name. The determinant has a clear geometrical meaning in the case of 2×2 and 3×3 matrices. In the former case the absolute value of the determinant is the area of a parallelogram formed with the matrix column vectors; in the latter case the absolute value of the determinant is the volume of the parallelepiped formed with the matrix column vectors. The determinant for $n \times n$ matrices is introduced as a suitable generalization of these properties. In this Section we present the determinant for 2×2 and 3×3 matrices and we study their main properties. We then present the definition of determinant for $n \times n$ matrices and we mention without proof its main properties.

3.1.1. **Determinant of 2×2 matrices.** We start introducing the determinant as a scalar-valued function on 2×2 matrices.

Definition 3.1.1. *The **determinant** of a 2×2 matrix $\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ is the value of the function $\det : \mathbb{F}^{2,2} \to \mathbb{F}$ given by $\det(\mathsf{A}) = A_{11}A_{22} - A_{12}A_{21}$.*

Depending on the context we will use for the determinant any of the following notations,

$$\det(\mathsf{A}) = |\mathsf{A}| = \Delta = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}.$$

For example, $\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = A_{11}A_{22} - A_{12}A_{21}$.

EXAMPLE 3.1.1: The value of the determinant can be any real number, as the following three cases show:

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 4 - 6 = -2, \qquad \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} = 8 - 3 = 5, \qquad \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 4 - 4 = 0.$$

◁

We have seen in Theorem 2.4.2 in Section 2.4 that a 2×2 matrix is invertible iff its determinant is non-zero. We now see that there is an interesting geometrical interpretation of this property.

Theorem 3.1.2. *Given a matrix $\mathsf{A} = [\mathsf{A}_1, \mathsf{A}_2] \in \mathbb{R}^{2,2}$, the absolute value of its determinant, $|\det(\mathsf{A})|$, is the area of the parallelogram formed by the vectors A_1, A_2.*

Proof of Theorem 3.1.2: Denote the matrix A as follows

$$\mathsf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \Rightarrow \quad \mathsf{A}_1 = \begin{bmatrix} a \\ c \end{bmatrix}, \quad \mathsf{A}_2 = \begin{bmatrix} b \\ d \end{bmatrix}.$$

The case where all the coefficients in matrix A are positive is shown in Fig. 27. We can see in this Figure that the area of the parallelogram formed by the vectors A_1 and A_2 is related to the area of the rectangle with sides b and c. More precisely, the area of the parallelogram is equal to the area of the rectangle minus the area of the two triangles marked with a "$-$" sign in Fig. 27 plus the area of the triangle marked with "$+$" sign. Denoting the area of the parallelogram by A_p, we obtain the following equation,

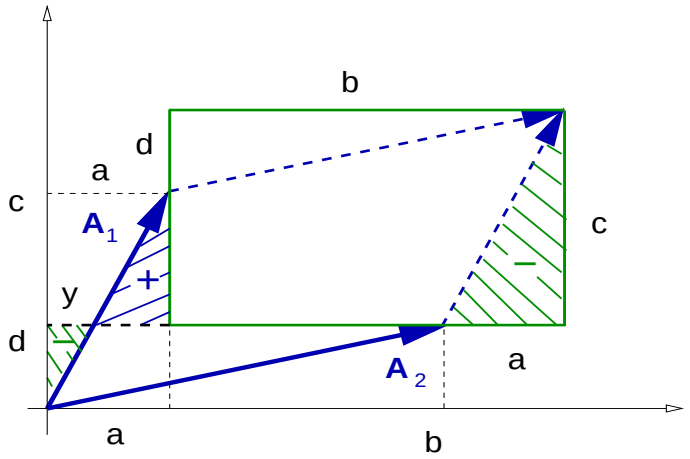

FIGURE 27. The geometrical meaning of the determinant of a 2×2 matrix is that its absolute value is the area of the parallelogram formed by the matrix column vectors.

$$\begin{aligned} A_p &= bc - \frac{ac}{2} - \frac{yd}{2} + \frac{(a-y)(c-d)}{2} \\ &= bc - \frac{ac}{2} - \frac{yd}{2} + \frac{ac}{2} - \frac{ad}{2} - \frac{yc}{2} + \frac{yd}{2} \\ &= bc - \frac{ad}{2} - \frac{yc}{2}. \end{aligned}$$

Similar triangles implies that

$$\frac{y}{d} = \frac{a}{c} \quad \Rightarrow \quad yc = ad.$$

Introducing this relation in the equation above it we obtain

$$A_p = bc - \frac{ad}{2} - \frac{ad}{2} \quad \Rightarrow \quad A_p = ad - bc = |\det(\mathsf{A})|.$$

We consider only this case in our proof. The remaining cases can be studied in a similar way. This establishes the Theorem. □

EXAMPLE 3.1.2: Find the area of the parallelogram formed by $\mathsf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathsf{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

SOLUTION: If we consider the matrix $\mathsf{A} = [\mathsf{a}, \mathsf{b}] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, then the area A of the parallelogram formed by the vectors a, b is $A = |\det(\mathsf{A})|$, that is,

$$\det(\mathsf{A}) = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = 1 - 4 = -3 \quad \Rightarrow \quad A = 3.$$

◁

We have seen that the determinant is a scalar-valued function on the space of 2×2 matrices, whose absolute value is the area of the parallelogram formed by the matrix column vectors . This relation of the determinant with areas on the plane is at the origin of the following properties.

Theorem 3.1.3. *For every vectors* a_1, a_2, $\mathsf{b} \in \mathbb{F}^2$ *and scalar* $k \in \mathbb{F}$, *the determinant function* $\det : \mathbb{F}^{2,2} \to \mathbb{F}$ *satisfies:*

(a) $\det\big([\mathsf{a}_1, \mathsf{a}_2]\big) = -\det\big([\mathsf{a}_2, \mathsf{a}_1]\big)$;
(b) $\det\big([k\,\mathsf{a}_1, \mathsf{a}_2]\big) = \det\big([\mathsf{a}_1, k\,\mathsf{a}_2]\big) = k\,\det\big([\mathsf{a}_1, \mathsf{a}_2]\big)$;
(c) $\det\big([\mathsf{a}_1, k\,\mathsf{a}_1]\big) = 0$;
(d) $\det\big([(\mathsf{a}_1 + \mathsf{b}), \mathsf{a}_2]\big) = \det\big([\mathsf{a}_1, \mathsf{a}_2]\big) + \det\big([\mathsf{b}, \mathsf{a}_2]\big)$.

Proof of Theorem 3.1.3: Introduce the notation

$$\mathsf{a}_1 = \begin{bmatrix} a \\ c \end{bmatrix}, \quad \mathsf{a}_2 = \begin{bmatrix} b \\ d \end{bmatrix}, \quad \Rightarrow \quad [\mathsf{a}_1, \mathsf{a}_2] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Part (a):

$$\left.\begin{aligned} \det\big([\mathsf{a}_1, \mathsf{a}_2]\big) &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \\ \det\big([\mathsf{a}_2, \mathsf{a}_1]\big) &= \begin{vmatrix} b & a \\ d & c \end{vmatrix} = bc - ad, \end{aligned}\right\} \quad \Rightarrow \quad \det\big([\mathsf{a}_1, \mathsf{a}_2]\big) = -\det\big([\mathsf{a}_2, \mathsf{a}_1]\big).$$

Part (b):

$$\det\big([k\,\mathsf{a}_1, \mathsf{a}_2]\big) = \begin{vmatrix} ka & b \\ kc & d \end{vmatrix} = k(ad - bc) = k \begin{vmatrix} a & b \\ c & d \end{vmatrix} = k\,\det\big([\mathsf{a}_1, \mathsf{a}_2]\big),$$

$$\det\big([\mathsf{a}_1, k\,\mathsf{a}_2]\big) = \begin{vmatrix} a & kb \\ c & kd \end{vmatrix} = k(ad - bc) = k \begin{vmatrix} a & b \\ c & d \end{vmatrix} = k\,\det\big([\mathsf{a}_1, \mathsf{a}_2]\big).$$

Part (c):

$$\det\big([\mathsf{a}_1, k\,\mathsf{a}_1]\big) = \det\big([k\,\mathsf{a}_1, \mathsf{a}_1]\big) = -\det\big([\mathsf{a}_1, k\,\mathsf{a}_1]\big) \quad \Rightarrow \quad \det\big([\mathsf{a}_1, k\,\mathsf{a}_1]\big) = 0.$$

Part (d): Introduce the notation $\mathsf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. Then,

$$\begin{aligned} \det\big([(\mathsf{a}_1 + \mathsf{b}), \mathsf{a}_2]\big) &= \begin{vmatrix} (a + b_1) & b \\ (c + b_2) & d \end{vmatrix} \\ &= (a + b_1)d - (c + b_2)b \\ &= (ad - cb) + (b_1 d - b_2 b) \\ &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} b_1 & b \\ b_2 & d \end{vmatrix} \\ &= \det\big([\mathsf{a}_1, \mathsf{a}_2]\big) + \det\big([\mathsf{b}, \mathsf{a}_2]\big). \end{aligned}$$

This establishes the Theorem. □

Moreover, the determinant function also satisfies the following further properties.

Theorem 3.1.4. *Let* A, $\mathsf{B} \in \mathbb{F}^{2,2}$. *Then it holds:*

(a) Matrix A *is invertible iff* $\det(\mathsf{A}) \neq 0$;
(b) $\det(\mathsf{A}) = \det(\mathsf{A}^T)$;
(c) $\det(\mathsf{AB}) = \det(\mathsf{A})\,\det(\mathsf{B})$.
(d) If matrix A *is invertible, then* $\det(\mathsf{A}^{-1}) = \dfrac{1}{\det(\mathsf{A})}$.

Proof of Theorem 3.1.4:

Part (a): It has been proven in Theorem 2.4.2.

Part (b): Denote $\mathsf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then holds

$$\det(\mathsf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = \det\big(\mathsf{A}^T\big).$$

Part (c): Denoting $\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and $\mathsf{B} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ we obtain,

$$\mathsf{AB} = \begin{bmatrix} (A_{11}B_{11} + A_{12}B_{21}) & (A_{11}B_{12} + A_{12}B_{22}) \\ (A_{21}B_{11} + A_{22}B_{21}) & (A_{21}B_{12} + A_{22}B_{22}) \end{bmatrix}.$$

Therefore,

$$\begin{aligned}
\det(\mathsf{AB}) &= (A_{11}B_{11} + A_{12}B_{21})(A_{21}B_{12} + A_{22}B_{22}) - (A_{11}B_{12} + A_{12}B_{22})(A_{21}B_{11} + A_{22}B_{21}) \\
&= A_{11}B_{11}A_{21}B_{12} + A_{11}B_{11}A_{22}B_{22} + A_{12}B_{21}A_{21}B_{12} + A_{12}B_{21}A_{22}B_{22} \\
&\quad - A_{21}B_{11}A_{11}B_{12} - A_{21}B_{11}A_{12}B_{22} - A_{22}B_{21}A_{11}B_{12} - A_{22}B_{21}A_{12}B_{22} \\
&= A_{11}B_{11}A_{22}B_{22} + A_{12}B_{21}A_{21}B_{12} - A_{21}B_{11}A_{12}B_{22} - A_{22}B_{21}A_{11}B_{12} \\
&= (A_{11}A_{22} - A_{21}A_{12})B_{11}B_{22} - (A_{11}A_{22} - A_{12}A_{21})B_{12}B_{21} \\
&= (A_{11}A_{22} - A_{21}A_{12})\,(B_{11}B_{22} - B_{12}B_{21}) \\
&= \det(\mathsf{A})\,\det(\mathsf{B}).
\end{aligned}$$

Part (d): Since $\mathsf{A}(\mathsf{A}^{-1}) = \mathsf{I}_2$, we obtain

$$1 = \det(\mathsf{I}_2) = \det\big(\mathsf{A}(\mathsf{A}^{-1})\big) = \det(\mathsf{A})\,\det\big(\mathsf{A}^{-1}\big) \quad \Rightarrow \quad \det\big(\mathsf{A}^{-1}\big) = \frac{1}{\det(\mathsf{A})}.$$

This establishes the Theorem. □

We finally mention that $\det(\mathsf{A}) = \det(\mathsf{A}^T)$ implies that Theorem 3.1.3 can be generalized from properties involving columns of the matrix into properties involving rows of the matrix. Introduce the following notation that generalizes column vectors to include row vectors:

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \big[\mathsf{a}_{:1},\, \mathsf{a}_{:2}\big] = \begin{bmatrix} \mathsf{a}_{1:} \\ \mathsf{a}_{2:} \end{bmatrix},$$

where

$$\mathsf{a}_{:1} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}, \quad \mathsf{a}_{:2} = \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}, \quad \mathsf{a}_{1:} = \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}, \quad \mathsf{a}_{2:} = \begin{bmatrix} A_{21} & A_{22} \end{bmatrix}.$$

The first two vectors above are the usual column vectors of matrix A, and the last two are its row vectors. When working with both, column vectors and row vectors, we use the notation above; when working only with column vectors, we drop the colon and we denote them, for example, as $\mathsf{A}_{:1} = \mathsf{A}_1$. Using this notation, it is simple to verify the following result.

Theorem 3.1.5. *For all column vectors $\mathsf{a}_{1:}^T, \mathsf{a}_{2:}^T, \mathsf{b}_{1:}^T \in \mathbb{F}^2$ and scalars $k \in \mathbb{F}$, the determinant function $\det : \mathbb{F}^{2,2} \to \mathbb{F}$ satisfies:*

(a) $\det\left(\begin{bmatrix} \mathsf{a}_{1:} \\ \mathsf{a}_{2:} \end{bmatrix}\right) = -\det\left(\begin{bmatrix} \mathsf{a}_{2:} \\ \mathsf{a}_{1:} \end{bmatrix}\right)$;

(b) $\det\left(\begin{bmatrix} k\,\mathsf{a}_{1:} \\ \mathsf{a}_{2:} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathsf{a}_{1:} \\ k\,\mathsf{a}_{2:} \end{bmatrix}\right) = k\,\det\left(\begin{bmatrix} \mathsf{a}_{1:} \\ \mathsf{a}_{2:} \end{bmatrix}\right)$;

(c) $\det\left(\begin{bmatrix} \mathsf{a}_{1:} \\ k\,\mathsf{a}_{1:} \end{bmatrix}\right) = 0$;

(d) $\det\left(\begin{bmatrix} (\mathsf{a}_{1:} + \mathsf{b}_{1:}) \\ \mathsf{a}_{2:} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathsf{a}_{1:} \\ \mathsf{a}_{2:} \end{bmatrix}\right) + \det\left(\begin{bmatrix} \mathsf{b}_{1:} \\ \mathsf{a}_{2:} \end{bmatrix}\right)$.

The proof is left as an exercise.

3.1.2. **Determinant of 3×3 matrices.** The determinant for 3×3 matrices is defined recursively in terms of 2×2 determinants.

Definition 3.1.6. *The* ***determinant*** *on 3×3 matrices is the function* $\det : \mathbb{F}^{3,3} \to \mathbb{F}$,

$$\det(\mathsf{A}) = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = A_{11} \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12} \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13} \begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}.$$

We see that the determinant of a 3×3 matrix is computed as a linear combination of the determinants of 2×2 matrices,

$$\det(\mathsf{A}) = A_{11}\,\det\big(\mathring{\mathsf{A}}_{11}\big) - A_{12}\,\det\big(\mathring{\mathsf{A}}_{12}\big) + A_{13}\,\det\big(\mathring{\mathsf{A}}_{13}\big),$$

where we introduced the 2×2 matrices

$$\mathring{\mathsf{A}}_{11} = \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}, \quad \mathring{\mathsf{A}}_{12} = \begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix}, \quad \mathring{\mathsf{A}}_{13} = \begin{bmatrix} A_{21} & A_{21} \\ A_{31} & A_{31} \end{bmatrix};$$

These matrices are three of the nine matrices called the minors of matrix A.

Definition 3.1.7. *The 2×2 matrix $\mathring{\mathsf{A}}_{ij}$ is called the (i,j)-**minor** of a 3×3 matrix A iff $\mathring{\mathsf{A}}_{ij}$ is obtained from A by removing the row i and the column j. The (i,j)-**cofactor** of matrix A is the number $C_{ij} = (-1)^{(i+j)} \det\big(\mathring{\mathsf{A}}_{ij}\big)$.*

These definitions are common in the literature, and they allow to write down a compact expression for the determinant of a 3×3 matrix. Indeed, the expression in Definition 3.1.6 now has the form

$$\det(\mathsf{A}) = A_{11}C_{11} + A_{12}C_{12} + A_{13}C_{13}.$$

Of course, all the complexity in the definition of determinant is now hidden in the definition of the cofactors. The latter expression is not simpler than the one in Definition 3.1.6, it only looks simpler.

EXAMPLE 3.1.3: Find the determinant of the matrix $\mathsf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}$.

SOLUTION: We use the definition above, that is,

$$\begin{aligned} \begin{vmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{vmatrix} &= 1 \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} - 3 \begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} + (-1) \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} \\ &= (1-2) - 3(2-3) - (4-3) \\ &= 1 + 3 - 1 \\ &= 1. \end{aligned}$$

We conclude that $\det(\mathsf{A}) = 1$. ◁

As in the 2×2 case, the determinant of a real-valued 3×3 matrix can be any real number, positive, negative of zero. The absolute value of the determinant has the geometrical meaning, see Fig. 28.

Theorem 3.1.8. *The number $|\det(\mathsf{A})|$, the absolute value of the determinant of matrix $\mathsf{A} = [\mathsf{A}_{:1}, \mathsf{A}_{:2}, \mathsf{A}_{:3}]$, is the volume of the parallelepiped formed by the vectors $\mathsf{A}_{:1}$, $\mathsf{A}_{:2}$, $\mathsf{A}_{:3}$.*

We do not give a prove of this statement. See the references at the end of the Section.

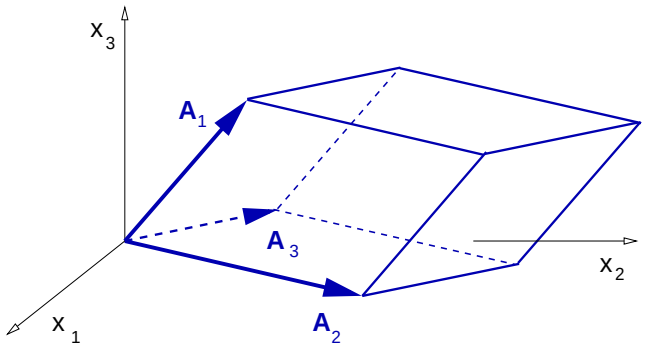

FIGURE 28. The geometrical meaning of the determinant of a 3×3 matrix is that its absolute value is the volume of the parallelepiped formed by the matrix column vectors.

EXAMPLE 3.1.4: Show that the determinant of an upper triangular or a lower triangular matrix is the product of its diagonal elements.

SOLUTION: Denote by A an upper-triangular matrix. Then, the definition of determinant of a 3×3 matrix implies

$$\det(\mathsf{A}) = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{vmatrix} = A_{11} \begin{vmatrix} A_{22} & A_{23} \\ 0 & A_{33} \end{vmatrix} - A_{12} \begin{vmatrix} 0 & A_{23} \\ 0 & A_{33} \end{vmatrix} + A_{13} \begin{vmatrix} 0 & A_{22} \\ 0 & 0 \end{vmatrix}.$$

We obtain that $\det(\mathsf{A}) = A_{11}A_{22}A_{33}$. Suppose now that A is lower-triangular. The definition of determinant of a 3×3 matrix implies

$$\det(\mathsf{A}) = \begin{vmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = A_{11} \begin{vmatrix} A_{22} & 0 \\ A_{32} & A_{33} \end{vmatrix} - 0 \begin{vmatrix} A_{21} & 0 \\ A_{31} & A_{33} \end{vmatrix} + 0 \begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}.$$

We obtain that $\det(\mathsf{A}) = A_{11}A_{22}A_{33}$. ◁

The determinant function on 3×3 satisfies a generalization of the properties proven for 2×2 matrices in Theorem 3.1.3.

Theorem 3.1.9. *For all vectors* a_1, a_2, a_3, $\mathsf{b} \in \mathbb{F}^3$ *and scalars* $k \in \mathbb{F}$, *the determinant function* $\det : \mathbb{F}^{3,3} \to \mathbb{F}$ *satisfies:*

(a) $\det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big) = -\det\big([\mathsf{a}_2, \mathsf{a}_1, \mathsf{a}_3]\big) = -\det\big([\mathsf{a}_1, \mathsf{a}_3, \mathsf{a}_2]\big)$;
(b) $\det\big([k\,\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big) = k\,\det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big)$;
(c) $\det\big([\mathsf{a}_1, k\,\mathsf{a}_1, \mathsf{a}_3]\big) = 0$;
(d) $\det\big([(\mathsf{a}_1 + \mathsf{b}), \mathsf{a}_2, \mathsf{a}_3]\big) = \det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big) + \det\big([\mathsf{b}, \mathsf{a}_2, \mathsf{a}_3]\big)$.

The proof of this Theorem is left as an exercise. The property (a) implies that all the remaining properties (b)-(d) also hold for all the column vectors. That is, from properties (a) and (b) one also shows

$$\det\big([k\,\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big) = \det\big([\mathsf{a}_1, k\,\mathsf{a}_2, \mathsf{a}_3]\big) = \det\big([\mathsf{a}_1, \mathsf{a}_2, k\,\mathsf{a}_3]\big) = k\,\det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big);$$

from properties (a) and (c) one also shows

$$\det\big([\mathsf{a}_1, \mathsf{a}_2, k\,\mathsf{a}_1]\big) = 0, \quad \det\big([\mathsf{a}_1, \mathsf{a}_2, k\,\mathsf{a}_2]\big) = 0;$$

and from properties (a) and (d) one also shows

$$\det\big([\mathsf{a}_1, (\mathsf{a}_2 + \mathsf{b}), \mathsf{a}_3]\big) = \det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big) + \det\big([\mathsf{a}_1, \mathsf{b}, \mathsf{a}_3]\big),$$
$$\det\big([\mathsf{a}_1, \mathsf{a}_2, (\mathsf{a}_3 + \mathsf{b})]\big) = \det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{a}_3]\big) + \det\big([\mathsf{a}_1, \mathsf{a}_2, \mathsf{b}]\big).$$

The following result is a generalization of Theorem 3.1.4.

Theorem 3.1.10. *Let* $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{3,3}$. *Then it holds:*

(a) *Matrix* A *is invertible iff* $\det(\mathsf{A}) \neq 0$;
(b) $\det(\mathsf{A}) = \det(\mathsf{A}^T)$;
(c) $\det(\mathsf{AB}) = \det(\mathsf{A})\,\det(\mathsf{B})$.
(d) *If matrix* A *is invertible, then* $\det(\mathsf{A}^{-1}) = \dfrac{1}{\det(\mathsf{A})}$.

The proof of this Theorem is left as an exercise. Like in the case of matrices 2×2, Theorem 3.1.9 can be generalized from properties involving column vector to properties involving row vectors.

Theorem 3.1.11. *For all vectors* $\mathsf{a}_{1:}^T, \mathsf{a}_{2:}^T, \mathsf{a}_{3:}^T, \mathsf{b}_{1:}^T \in \mathbb{F}^3$ *and scalars* $k \in \mathbb{F}$, *the determinant function* $\det : \mathbb{F}^{3,3} \to \mathbb{F}$ *satisfies:*

(a) $\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = -\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{2:}\\ \mathsf{a}_{1:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = -\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{3:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{1:}\end{bmatrix}\end{pmatrix} = -\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{3:}\\ \mathsf{a}_{2:}\end{bmatrix}\end{pmatrix}$;

(b) $\det\begin{pmatrix}\begin{bmatrix}k\,\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = k\,\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix}$;

(c) $\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ k\,\mathsf{a}_{1:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = 0$;

(d) $\det\begin{pmatrix}\begin{bmatrix}(\mathsf{a}_{1:} + \mathsf{b}_{1:})\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} + \det\begin{pmatrix}\begin{bmatrix}\mathsf{b}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix}$.

The proof is left as an exercise. The property (a) implies that all the remaining properties (b)-(d) also hold for all the row vectors. That is, from properties (a) and (b) one also shows

$$\det\begin{pmatrix}\begin{bmatrix}k\,\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ k\,\mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ k\,\mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = k\,\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix};$$

from properties (a) and (c) one also shows

$$\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ k\,\mathsf{a}_{1:}\end{bmatrix}\end{pmatrix} = 0; \quad \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ k\,\mathsf{a}_{2:}\\ \mathsf{a}_{2:}\end{bmatrix}\end{pmatrix} = 0;$$

from properties (a) and (d) one also shows

$$\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ (\mathsf{a}_{2:} + \mathsf{b}_{2:})\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} = \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} + \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{b}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix},$$
$$\det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ (\mathsf{a}_{3:} + \mathsf{b}_{3:})\end{bmatrix}\end{pmatrix} = \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{a}_{3:}\end{bmatrix}\end{pmatrix} + \det\begin{pmatrix}\begin{bmatrix}\mathsf{a}_{1:}\\ \mathsf{a}_{2:}\\ \mathsf{b}_{3:}\end{bmatrix}\end{pmatrix}.$$

3.1.3. **Determinant of $n \times n$ matrices.** The notion of determinant can be generalized to $n \times n$ matrices with $n \geqslant 1$. One way to find an appropriate generalization is to realize that the determinant of 2×2 and 3×3 matrices are related to the notion of areas and volumes, respectively. One then studies the properties of areas an volumes, like the ones given in Theorem 3.1.3 and 3.1.9. It can be shown that generalizations of these properties determine a unique function $\det : \mathbb{F}^{n,n} \to \mathbb{F}$. In this notes we only present the final result, the definition of determinant for $n \times n$ matrices. The reader is referred to the literature for a constructive proof of the function determinant.

Definition 3.1.12. *The **determinant** of a matrix $\mathsf{A} = [A_{ij}] \in \mathbb{F}^{n,n}$, with $n \geqslant 1$ and $i, j = 1, \cdots, n$, is the value of the function $\det : \mathbb{F}^{n,n} \to \mathbb{F}$ is given by*

$$\det(\mathsf{A}) = A_{11}\, C_{11} + \cdots + A_{1n}\, C_{1n}, \tag{3.1}$$

*where the number C_{ij}, called (i,j)**-cofactor** of matrix A, is the scalar given by*

$$C_{ij} = (-1)^{(i+j)} \det\big(\mathring{\mathsf{A}}_{ij}\big),$$

*and where the matrices $\mathring{\mathsf{A}}_{ij} \in \mathbb{F}^{(n-1),(n-1)}$, called the (i,j)**-minors** of A, are obtained from matrix A by eliminating the row i and the column j, which are highlighted below*

$$\mathring{\mathsf{A}}_{ij} = \begin{bmatrix} A_{11} & \cdots & \boldsymbol{A_{1j}} & \cdots & A_{1n} \\ \vdots & & \vdots & & \vdots \\ \boldsymbol{A_{i1}} & \cdots & \boldsymbol{A_{ij}} & \cdots & \boldsymbol{A_{in}} \\ \vdots & & \vdots & & \vdots \\ A_{n1} & \cdots & \boldsymbol{A_{nj}} & \cdots & A_{nn} \end{bmatrix}.$$

EXAMPLE 3.1.5: Use Eq. (3.1) to compute the determinant of a 2×2 matrix.

SOLUTION: Consider the matrix $\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. The four minor matrices are given by

$$\mathring{\mathsf{A}}_{11} = A_{22}, \quad \mathring{\mathsf{A}}_{12} = A_{21}, \quad \mathring{\mathsf{A}}_{21} = A_{12}, \quad \mathring{\mathsf{A}}_{22} = A_{11}.$$

This means, generalizing the notion of determinant to numbers by $\det(a) = a$, that the cofactors are

$$C_{11} = A_{22}, \quad C_{12} = -A_{21}, \quad C_{21} = -A_{12}, \quad C_{22} = A_{11}.$$

So, we obtain that

$$\det(\mathsf{A}) = A_{11}C_{11} + A_{12}C_{12} \quad \Leftrightarrow \quad \det(\mathsf{A}) = A_{11}A_{22} - A_{12}A_{21}.$$

Notice that we do not need to compute all four cofactors to compute the determinant of the matrix; just two of them are enough. ◁

EXAMPLE 3.1.6: Use Eq. (3.1) to compute the determinant of a 3×3 matrix.

SOLUTION: Consider the matrix

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}.$$

The nine minor matrices are given by

$$\mathring{\mathsf{A}}_{11} = \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} \qquad \mathring{\mathsf{A}}_{12} = \begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix} \qquad \mathring{\mathsf{A}}_{13} = \begin{bmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}$$
$$\mathring{\mathsf{A}}_{21} = \begin{bmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{bmatrix} \qquad \mathring{\mathsf{A}}_{22} = \begin{bmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{bmatrix} \qquad \mathring{\mathsf{A}}_{23} = \begin{bmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{bmatrix}$$
$$\mathring{\mathsf{A}}_{31} = \begin{bmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{bmatrix} \qquad \mathring{\mathsf{A}}_{32} = \begin{bmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{bmatrix} \qquad \mathring{\mathsf{A}}_{23} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Then, Def. 3.1.12 agrees with Def. 3.1.6, since

$$\det(\mathsf{A}) = A_{11}C_{11} + A_{12}C_{12} + A_{13}C_{13}$$

is equivalent to

$$\det(\mathsf{A}) = A_{11}\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13}\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}.$$

◁

We now state several results that summarize the main properties of the determinant of $n \times n$ matrices. We state the results without proof. The frost result says that the determinant of a matrix can be computed using expansions along any row or any column in the matrix.

Theorem 3.1.13. *The determinant of an $n \times n$ matrix $\mathsf{A} = [A_{ij}]$ can be computed expanding along any row or any column of matrix A, that is,*

$$\begin{aligned} \det(\mathsf{A}) &= A_{i1}C_{i1} + \cdots + A_{in}C_{in}, & i &= 1, \cdots, n, \\ &= A_{1j}C_{1j} + \cdots + A_{nj}C_{nj}, & j &= 1, \cdots, n. \end{aligned}$$

EXAMPLE 3.1.7: Show all possible expansions of the determinant of a 3×3 matrix $\mathsf{A} = [A_{ij}]$.

SOLUTION: The expansions along each of the three rows are the following:

$$\begin{aligned} \det(\mathsf{A}) &= A_{11}C_{11} + A_{12}C_{12} + A_{13}C_{13} \\ &= A_{21}C_{21} + A_{22}C_{22} + A_{23}C_{23} \\ &= A_{31}C_{31} + A_{32}C_{32} + A_{33}C_{33}; \end{aligned}$$

The expansions along each of the three columns are the following:

$$\begin{aligned} \det(\mathsf{A}) &= A_{11}C_{11} + A_{21}C_{21} + A_{31}C_{31} \\ &= A_{12}C_{12} + A_{22}C_{22} + A_{32}C_{32} \\ &= A_{13}C_{13} + A_{23}C_{23} + A_{33}C_{33}. \end{aligned}$$

◁

EXAMPLE 3.1.8: Compute the determinant of $\mathsf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}$ with an expansion by the third column.

SOLUTION: The expansion by the third column is the following,

$$\begin{vmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{vmatrix} = (-1)\begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} - (1)\begin{vmatrix} 1 & 3 \\ 3 & 2 \end{vmatrix} + (1)\begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix}$$
$$= -(4-3) - (2-9) + (1-6)$$
$$= -1 + 7 - 5$$
$$= 1,$$

That is, $\det(\mathsf{A}) = 1$. This result agrees with Example 3.1.3. ◁

The generalization of Theorem 3.1.9 to $n \times n$ matrices is the following.

Theorem 3.1.14. *For all vectors a_i, a_j,$\mathsf{b} \in \mathbb{F}^n$, with $i, j = 1, \cdots, n$, and for all scalars $k \in \mathbb{F}$, the determinant function $\det : \mathbb{F}^{n,n} \to \mathbb{F}$ satisfies:*

(a) $\det\big([\mathsf{a}_1, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_n]\big) = -\det\big([\mathsf{a}_1, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_n]\big)$;

(b) $\det\big([\mathsf{a}_1, \cdots, k\,\mathsf{a}_i, \cdots, \mathsf{a}_n]\big) = k\,\det\big([\mathsf{a}_1, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_n]\big)$;

(c) $\det\big([\mathsf{a}_1, \cdots, k\,\mathsf{a}_i, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_n]\big) = 0$;

(d) $\det\big([\mathsf{a}_1, \cdots, (\mathsf{a}_i + \mathsf{b}), \cdots, \mathsf{a}_n]\big) = \det\big([\mathsf{a}_1, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_n]\big) + \det\big([\mathsf{b}, \cdots, \mathsf{b}, \cdots, \mathsf{a}_n]\big)$.

The generalization of Theorem 3.1.10 to $n \times n$ matrices is the following.

Theorem 3.1.15. *Let $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{n,n}$. Then it holds:*

(a) Matrix A is invertible iff $\det(\mathsf{A}) \neq 0$;

(b) $\det(\mathsf{A}) = \det(\mathsf{A}^T)$;

(c) $\det(\overline{\mathsf{A}}) = \overline{\det(\mathsf{A})}$;

(d) $\det(\mathsf{AB}) = \det(\mathsf{A})\,\det(\mathsf{B})$.

(e) If matrix A is invertible, then $\det(\mathsf{A}^{-1}) = \dfrac{1}{\det(\mathsf{A})}$.

The proof of this Proposition is left as an exercise. Like in the case of matrices 2×2, Theorem 3.1.9 can be generalized from properties involving column vector to properties involving row vectors.

Theorem 3.1.16. *For all vectors $\mathsf{a}_{i:}^T$, $\mathsf{a}_{j:}^T$, $\mathsf{b}_{i:}^T \in \mathbb{F}^n$ and scalars $k \in \mathbb{F}$, the determinant function $\det : \mathbb{F}^{n,n} \to \mathbb{F}$ satisfies:*

$$\begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ \mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{j:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix} = -\begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ \mathsf{a}_{j:} \\ \vdots \\ \mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix}; \qquad \begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ k\,\mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix} = k \begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ \mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix}; \qquad \begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ k\,\mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix} = 0; \qquad \begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ (\mathsf{a}_{i:} + \mathsf{b}_{i:}) \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix} = \begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ \mathsf{a}_{i:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix} + \begin{vmatrix} \mathsf{a}_{1:} \\ \vdots \\ \mathsf{b}_{i:} \\ \vdots \\ \mathsf{a}_{n:} \end{vmatrix}.$$

3.1.4. **Exercises.**

3.1.1.- Find the determinant of matrices

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 6 & 2 & 1 \\ -2 & 2 & 1 \end{bmatrix}, \mathsf{B} = \begin{bmatrix} 2 & 1 & 1 \\ 6 & 0 & 1 \\ -2 & 0 & 1 \end{bmatrix}.$$

3.1.2.- Find the volume of the parallelepiped formed by the vectors

$$\mathsf{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}, \ \mathsf{x}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \ \mathsf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

3.1.3.- Find the determinants of the upper and lower triangular matrices

$$\mathsf{U} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}, \mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}.$$

3.1.4.- Find the determinant of the matrix

$$\mathsf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

3.1.5.- Give an example to show that in general

$$\det(\mathsf{A} + \mathsf{B}) \neq \det(\mathsf{A}) + \det(\mathsf{B}).$$

3.1.6.- If $\mathsf{A} \in \mathbb{F}^{n,n}$ express

$$\det(2\mathsf{A}), \quad \det(-\mathsf{A}), \quad \det(\mathsf{A}^2)$$

in terms of $\det(\mathsf{A})$ and n.

3.1.7.- Given matrices $\mathsf{A}, \mathsf{P} \in \mathbb{F}^{n,n}$, with P invertible, let $\mathsf{B} = \mathsf{P}^{-1}\mathsf{A}\,\mathsf{P}$. Prove that $\det(\mathsf{B}) = \det(\mathsf{A})$.

3.1.8.- Prove that for all matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ holds that $\det(\mathsf{A}^*) = \overline{\det(\mathsf{A})}$.

3.1.9.- Prove that for all $\mathsf{A} \in \mathbb{F}^{n,n}$ holds that $\det(\mathsf{A}^*\mathsf{A}) \geqslant 0$.

3.1.10.- Prove that for all $\mathsf{A} \in \mathbb{F}^{n,n}$ and all $k \in \mathbb{F}$ holds that $\det(k\,\mathsf{A}) = k^n \det(\mathsf{A})$.

3.1.11.- Prove that a skew-symmetric matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, with n odd, must satisfy that $\det(\mathsf{A}) = 0$.

3.1.12.- Let $\mathsf{A} \in \mathbb{F}^{n,n}$ be a matrix satisfying that $\mathsf{A}^T\mathsf{A} = \mathsf{I}_n$. Prove that

$$\det(\mathsf{A}) = \pm 1.$$

3.2. Applications

In the previous section we mentioned that the absolute value of the determinant is the area of a parallelogram formed by the column vectors of a 2×2 matrix, while it is the volume of the parallelepiped formed by the column vectors of a 3×3 matrix. There are several other applications of determinants. They determine whether a matrix $\mathsf{A}\in\mathbb{F}^{n,n}$ is invertible or not, since a matrix A is invertible iff $\det(\mathsf{A})\neq 0$. They determine whether a square system of linear equations has a unique solution for every source vector. The determinant is the key tool to find a formula for the inverse matrix, hence a formula for the solution components of a square linear system, called Cramer's rule. We discuss these to applications in more detail below.

3.2.1. **Inverse matrix formula.** The determinant of a matrix plays a crucial role in a formula for the inverse matrix. The existence of this formula does not change the fact that one can always compute the inverse matrix using Gauss operations. This formula for the inverse matrix is important though, since it explicitly shows how the inverse matrix depends on the coefficients of the original matrix. It thus shows how the inverse changes when the original matrix changes. The formula for the inverse matrix shows that the inverse matrix is a continuous function of the original matrix.

Theorem 3.2.1. *Given a matrix $\mathsf{A}=[A_{ij}]\in\mathbb{F}^{n,n}$, let $\mathsf{C}=[C_{ij}]$ be its cofactor matrix, where the cofactors are given by $C_{ij}=(-1)^{(i+j)}\det(\mathring{\mathsf{A}}_{ij})$, and matrix $\mathring{\mathsf{A}}_{ij}\in\mathbb{F}^{(n-1),(n-1)}$ is the (i,j)-minor of matrix A. If $\det(\mathsf{A})\neq 0$, then the inverse of matrix A is given by*

$$\mathsf{A}^{-1}=\frac{1}{\det(\mathsf{A})}\,\mathsf{C}^T,\qquad \textit{that is,}\qquad \big(\mathsf{A}^{-1}\big)_{ij}=\frac{1}{\det(\mathsf{A})}\,C_{ji}.$$

Proof of Theorem 3.2.1: Since $\det(\mathsf{A})\neq 0$ we know that matrix A is invertible. We only need to verify that the expression for A^{-1} given in Theorem 3.2.1 is correct. That is, we need to show that $\mathsf{C}^T\mathsf{A}=\det(\mathsf{A})\,\mathsf{I}_n=\mathsf{A}\,\mathsf{C}^T$. We start with the product

$$(\mathsf{C}^T\mathsf{A})_{ij}=\big(C_{1i}\boldsymbol{A}_{1j}+\cdots+C_{ni}\boldsymbol{A}_{nj}\big).$$

Notice that this component $(\mathsf{C}^T\mathsf{A})_{ij}$ is precisely the the expansion along the column i of the determinant of the following matrix: The matrix constructed from matrix A by placing the column vector $\mathsf{A}_{j:}$ in both columns i and j. In the particular case that $i<j$, this component $(\mathsf{C}^T\mathsf{A})_{ij}$ has the form

$$\big(C_{1i}\boldsymbol{A}_{1j}+\cdots+C_{ni}\boldsymbol{A}_{nj}\big)=\begin{vmatrix} A_{11} & \cdots & \boldsymbol{A}_{1j} & \cdots & A_{1j} & \cdots & A_{1n}\\ \vdots & & \vdots & & \vdots & & \vdots\\ A_{i1} & \cdots & \boldsymbol{A}_{ij} & \cdots & A_{ij} & \cdots & A_{in}\\ \vdots & & \vdots & & \vdots & & \vdots\\ A_{j1} & \cdots & \boldsymbol{A}_{jj} & \cdots & A_{jj} & \cdots & A_{jn}\\ \vdots & & \vdots & & \vdots & & \vdots\\ A_{n1} & \cdots & \boldsymbol{A}_{nj} & \cdots & A_{nj} & \cdots & A_{nn}\end{vmatrix},\quad i<j,$$

where we have highlighted in blue the column i which is occupied by the column vector $\mathsf{A}_{:j}$. The determinant on the right hand side vanishes, since for $i<j$ the matrix has the columns i and j repeated. A similar situation happens for $i>j$. So, we conclude that the non-diagonal elements of $(\mathsf{C}^T\mathsf{A})$ vanish. The diagonal elements are given by

$$(\mathsf{C}^T\mathsf{A})_{ii}=\big(C_{1i}\boldsymbol{A}_{1i}+\cdots+C_{ni}\boldsymbol{A}_{ni}\big)=\det(\mathsf{A}),$$

the determinant of A expanded along the i-th column. This shows that $\mathsf{C}^T\mathsf{A}=\det(\mathsf{A})\,\mathsf{I}_n$. A similar analysis shows that $\mathsf{A}\,\mathsf{C}^T=\det(\mathsf{A})\,\mathsf{I}_n$. This establishes the Theorem. □

EXAMPLE 3.2.1: Use the formula in Theorem 3.2.1 to find the inverse of matrix

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}.$$

SOLUTION: We already know from Example 3.1.3 that $\det(A) = 1$. We now need to compute all the cofactors:

$$C_{11} = \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} = -1; \qquad C_{12} = -\begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} = 1; \qquad C_{13} = \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} = 1;$$

$$C_{21} = -\begin{vmatrix} 3 & -1 \\ 2 & 1 \end{vmatrix} = -5; \qquad C_{22} = \begin{vmatrix} 1 & -1 \\ 3 & 1 \end{vmatrix} = 4; \qquad C_{23} = -\begin{vmatrix} 1 & 3 \\ 3 & 2 \end{vmatrix} = 7;$$

$$C_{31} = \begin{vmatrix} 3 & -1 \\ 1 & 1 \end{vmatrix} = 4; \qquad C_{32} = -\begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} = -3; \qquad C_{33} = \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} = -5.$$

Therefore, the cofactor matrix is given by

$$C = \begin{bmatrix} -1 & 1 & 1 \\ -5 & 4 & 7 \\ 4 & -3 & -5 \end{bmatrix},$$

and the formula $A^{-1} = C^T / \det(A)$ together with $\det(A) = 1$ imply that

$$A^{-1} = \begin{bmatrix} -1 & -5 & 4 \\ 1 & 4 & -3 \\ 1 & 7 & -5 \end{bmatrix}.$$

◁

The formula in Theorem 3.2.1 is useful to compute individual components of the inverse of a matrix.

EXAMPLE 3.2.2: Find the coefficients $\left(A^{-1}\right)_{12}$ and $\left(A^{-1}\right)_{32}$ of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & 3 \end{bmatrix}.$$

SOLUTION: We first need to compute

$$\det(A) = (1)\begin{vmatrix} -1 & 2 \\ 1 & 3 \end{vmatrix} - (1)\begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} + (1)\begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = -5 - 1 + 2 \quad \Rightarrow \quad \det(A) = -4.$$

The formula in Theorem 3.2.1 implies that

$$\left(A^{-1}\right)_{12} = \frac{1}{-4} C_{21}, \quad C_{21} = (-1)\begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} = -2, \quad \Rightarrow \quad \left(A^{-1}\right)_{12} = \frac{1}{2}.$$

$$\left(A^{-1}\right)_{32} = \frac{1}{-4} C_{23}, \quad C_{23} = (-1)\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0, \quad \Rightarrow \quad \left(A^{-1}\right)_{32} = 0.$$

◁

3.2.2. **Cramer's rule.** We know that given an invertible matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, a system of linear equations $\mathsf{Ax} = \mathsf{b}$ has a unique solution x for every source vector $\mathsf{b} \in \mathbb{F}^n$. This solution can be written in terms of the inverse matrix A^{-1} as $\mathsf{x} = \mathsf{A}^{-1}\mathsf{b}$. The formula for the inverse matrix given in Theorem 3.2.1 provides an explicit expression for the solution x. The result is known as Cramer's rule, and it is summarized below.

Theorem 3.2.2. *If the matrix $\mathsf{A} = [\mathsf{A}_1, \cdots, \mathsf{A}_n] \in \mathbb{F}^{n,n}$ is invertible, then the system of linear equations $\mathsf{Ax} = \mathsf{b}$ has a unique solution $\mathsf{x} = [x_i]$ for every $\mathsf{b} \in \mathbb{F}^n$ given by*

$$x_i = \frac{\det\big(\mathsf{A}_i(\mathsf{b})\big)}{\det(\mathsf{A})},$$

with matrix $\mathsf{A}_i(\mathsf{b}) = [\mathsf{A}_1, \cdots, \mathsf{b}, \cdots, \mathsf{A}_n]$ where vector b is placed in the i-th column.

EXAMPLE 3.2.3: Use Cramer's rule to find the solution x of the linear system $\mathsf{Ax} = \mathsf{b}$, where

$$\mathsf{A} = \begin{bmatrix} 3 & -2 \\ -5 & 6 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} 7 \\ -5 \end{bmatrix}.$$

SOLUTION: We first need to compute the determinant of A, that is,

$$\det(\mathsf{A}) = 18 - 10 \quad \Rightarrow \quad \det(\mathsf{A}) = 8.$$

Then, we need to find the matrices $\mathsf{A}_1(\mathsf{b})$ and $\mathsf{A}_2(\mathsf{b})$, given by

$$\mathsf{A}_1(\mathsf{b}) = [\mathsf{b}, \mathsf{A}_2] = \begin{bmatrix} 7 & -2 \\ -5 & 6 \end{bmatrix}, \qquad \mathsf{A}_2(\mathsf{b}) = [\mathsf{A}_1, \mathsf{b}] = \begin{bmatrix} 3 & 7 \\ -5 & -5 \end{bmatrix}.$$

We finally compute their determinants,

$$\det(\mathsf{A}_1(\mathsf{b})) = 42 - 10 = 32, \qquad \det(\mathsf{A}_2(\mathsf{b})) = -15 + 35 = 20.$$

So the solution is,

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad x_1 = \frac{32}{8} = 4, \qquad x_2 = \frac{20}{8} = \frac{5}{2}, \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 4 \\ 5/2 \end{bmatrix}.$$

◁

Proof of Theorem 3.2.2: Since matrix A is invertible, the solution to the linear system is $\mathsf{x} = \mathsf{A}^{-1}\mathsf{b}$. Using the formula for the inverse matrix given in Theorem 3.2.1, the solution $\mathsf{x} = [x_i]$ can be written as

$$\mathsf{x} = \frac{1}{\det(\mathsf{A})}\,\mathsf{C}^T\mathsf{b} \quad \Rightarrow \quad x_i = \frac{1}{\det(\mathsf{A})}\sum_{j=1}^{n} C_{ji}b_j.$$

Notice that the sum in the last equation above is precisely the expansion along the column i of the determinant of the matrix $\mathsf{A}_j(\mathsf{b})$, that is,

$$\det(\mathsf{A}_i(\mathsf{b})) = \begin{vmatrix} A_{11} & \cdots & b_1 & \cdots & A_{1n} \\ \vdots & & \vdots & & \vdots \\ A_{n1} & \cdots & b_n & \cdots & A_{nn} \end{vmatrix} = b_1C_{1i} + \cdots + b_nC_{ni} = \sum_{j=1}^{n} b_jC_{ji},$$

where the vector b replaces the vector A_i in the column i of matrix A. So, we conclude that

$$x_i = \frac{\det(\mathsf{A}_i(\mathsf{b}))}{\det(\mathsf{A})}.$$

This establishes the Theorem. □

3.2.3. **Determinants and Gauss operations.** Gauss elimination operations can be used to compute the determinant of a matrix. If a Gauss operation on matrix A produces matrix B, then $\det(\mathsf{A})$ is related to $\det(\mathsf{B})$ in a precise way. Here is the relation.

Theorem 3.2.3. *Let $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{m,n}$ be matrices related by a Gauss operation, that is, $\mathsf{A} \to \mathsf{B}$ by a Gauss operation. Then, the following statements hold:*

(a) If matrix B is the result of adding to a row in A a multiple of another row in A, then

$$\det(\mathsf{A}) = \det(\mathsf{B});$$

(b) If matrix B is the result of interchanging two rows in A, then

$$\det(\mathsf{A}) = -\det(\mathsf{B});$$

(c) If matrix B is the result of the multiplication of one row in A by a scalar $\dfrac{1}{k} \in \mathbb{F}$, then

$$\det(\mathsf{A}) = k\,\det(\mathsf{B}).$$

Proof of Theorem 3.2.3: We use the row vector notation for matrices A and B,

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{n:} \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} \mathsf{B}_{1:} \\ \vdots \\ \mathsf{B}_{n:} \end{bmatrix}.$$

Part (a): Matrix B results from multiplying row j of A by k and adding that to row i,

$$\mathsf{B} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} + k\mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{bmatrix} \quad \Rightarrow \quad \det(\mathsf{B}) = \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} + k\mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} + k \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = \det(\mathsf{A}).$$

Part (b): Matrix B is the result of the interchange of rows i and j in matrix A, that is,

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{bmatrix} \to \mathsf{B} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{n:} \end{bmatrix} \quad \Rightarrow \quad \det(\mathsf{B}) = \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = - \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{j:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = -\det(\mathsf{A}).$$

Part (c): Matrix B is the result of the multiplication of row i of A by $\dfrac{1}{k} \in \mathbb{F}$, that is,

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{n:} \end{bmatrix} \to \mathsf{B} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \frac{1}{k}\,\mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{n:} \end{bmatrix} \quad \Rightarrow \quad \det(\mathsf{B}) = \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \frac{1}{k}\,\mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = \frac{1}{k} \begin{vmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{i:} \\ \vdots \\ \mathsf{A}_{n:} \end{vmatrix} = \frac{1}{k}\,\det(\mathsf{A}).$$

This establishes the Theorem. □

EXAMPLE 3.2.4: Use Gauss operations to transform matrix A below into upper-triangular form, and use that calculation to find det(A), where,

$$\mathsf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}.$$

SOLUTION: We perform Gauss operations to transform A into upper-triangular form:

$$\det(\mathsf{A}) = \begin{vmatrix} 1 & 3 & -1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & -1 \\ 0 & -5 & 3 \\ 0 & -7 & 4 \end{vmatrix} = (-5) \begin{vmatrix} 1 & 3 & -1 \\ 0 & 1 & -3/5 \\ 0 & -7 & 4 \end{vmatrix} = (-5) \begin{vmatrix} 1 & 3 & -1 \\ 0 & 1 & -3/5 \\ 0 & 0 & -1/5 \end{vmatrix}.$$

We only need to reduce matrix A into upper-triangular form, not into reduced echelon form, since the determinant of an upper-triangular matrix is simple enough to find, just the product of its diagonal elements. In our case we find that

$$\det(\mathsf{A}) = (-5) \begin{vmatrix} 1 & 3 & -1 \\ 0 & 1 & -3/5 \\ 0 & 0 & -1/5 \end{vmatrix} = (-5)(1)(1)\Big(-\frac{1}{5}\Big) \quad \Rightarrow \quad \det(\mathsf{A}) = 1.$$

◁

3.2.4. **Exercises.**

3.2.1.- Use determinants to compute A^{-1}, where

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 6 & 2 & 1 \\ -2 & 2 & 1 \end{bmatrix}.$$

3.2.2.- Find the coefficients $(\mathsf{A}^{-1})_{12}$ and $(\mathsf{A}^{-1})_{32}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 5 & 7 \\ 2 & 1 & 0 \\ 4 & 1 & 3 \end{bmatrix}.$$

3.2.3.- Use Gauss operations to reduce matrices A and B below to upper triangular form and evaluate their determinant, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \\ 1 & 4 & 4 \end{bmatrix}, \mathsf{B} = \begin{bmatrix} 1 & 3 & 5 \\ -1 & 4 & 2 \\ 3 & -2 & 4 \end{bmatrix}.$$

3.2.4.- Use Gauss operations to prove the formula

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$$

3.2.5.- Use the $\det(\mathsf{A})$ to find the values of the constant k such that the system $\mathsf{A}\mathsf{x} = \mathsf{b}$ has a unique solution for every source vector b, where

$$\mathsf{A} = \begin{bmatrix} 1 & k & 0 \\ 0 & 1 & -1 \\ k & 0 & 1 \end{bmatrix}.$$

3.2.6.- Use Cramer's rule to find the solution to the linear system

$$a\,x_1 + b\,x_2 = 1$$
$$c\,x_1 + d\,x_2 = 0.$$

where a, b, c, $d \in \mathbb{R}$.

3.2.7.- Use Cramer's rule to find the solution to the linear system

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Chapter 4. Vector spaces

4.1. Spaces and subspaces

A vector space is any set where the linear combination operation is defined on its elements. In a vector space, also called a linear space, the elements are not important. The actual elements that constitute the vector space are left unspecified, only the relation among them is determined. An example of a vector space is the set $\mathbb{F}^n$ of n-vectors with the operation of linear combination studied in Chapter 1. Another example is the set $\mathbb{F}^{m,n}$ of all $m \times n$ matrices with the operation of linear combination studied in Chapter 2. We now define a vector space and comment its main properties. A subspace is introduced later on as a smaller vector space inside the original one. We end this Section with the concept of span of a set of vectors, which is a way to construct a subspace from any subset in a vector space.

Definition 4.1.1. *A set V is a **vector space** over the scalar field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ iff there are two operations defined on V, called vector addition and scalar multiplication with the following properties: For all $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w} \in V$ the vector addition satisfies*

(A1) $\boldsymbol{u} + \boldsymbol{v} \in V$, *(closure of V);*
(A2) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$, *(commutativity);*
(A3) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$, *(associativity);*
(A4) $\exists\ \boldsymbol{0} \in V \ :\ \boldsymbol{0} + \boldsymbol{u} = \boldsymbol{u} \quad \forall\ \boldsymbol{u} \in V$, *(existence of a neutral element);*
(A5) $\forall\ \boldsymbol{u} \in V \quad \exists\ (-\boldsymbol{u}) \in V \ :\ \boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$, *(existence of an opposite element);*

furthermore, for all a, $b \in \mathbb{F}$ the scalar multiplication satisfies

(M1) $a\boldsymbol{u} \in V$, *(closure of V);*
(M2) $1\,\boldsymbol{u} = \boldsymbol{u}$, *(neutral element);*
(M3) $a(b\boldsymbol{u}) = (ab)\boldsymbol{u}$, *(associativity);*
(M4) $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$, *(distributivity);*
(M5) $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$, *(distributivity).*

The definition of a vector space does not specify the elements of the set V, it only determines the properties of the vector addition and scalar multiplication operations. We use the convention that elements in a vector space, called vectors, are represented in boldface. Nevertheless, we allow several exceptions, the first two of them are given in Examples 4.1.1 and 4.1.2. We now present several examples of vector spaces.

Example 4.1.1: A vector space is the set $\mathbb{F}^n$ of n-vectors $\mathsf{u} = [u_i]$ with components $u_i \in \mathbb{F}$ and the operations of column vector addition and scalar multiplication given by

$$[u_i] + [v_i] = [u_i + v_i], \qquad a[u_i] = [au_i].$$

This space of column vectors was introduced in Chapter 1. Elements in these vector spaces are not represented in boldface, instead we keep the previous sanserif font, $\mathsf{u} \in \mathbb{F}^n$. The reason for this notation will be clear in Sect. 4.4. ◁

Example 4.1.2: A vector space is the set $\mathbb{F}^{m,n}$ of $m \times n$ matrices $\mathsf{A} = [A_{ij}]$ with matrix coefficients $A_{ij} \in \mathbb{F}$ and the operations of addition and scalar multiplication given by

$$\big[A_{ij}\big] + \big[B_{ij}\big] = \big[A_{ij} + B_{ij}\big], \qquad a\big[A_{ij}\big] = \big[aA_{ij}\big],$$

These operations were introduced in Chapter 2. As in the previous example, elements in these vector spaces are not represented in boldface, instead we keep the previous capital sanserif font, $\mathsf{A} \in \mathbb{F}^{m,n}$. The reason for this notation will be clear in Sect. 4.4. ◁

EXAMPLE 4.1.3: Let $\mathbb{P}_n(U)$ be the set of all polynomials having degree $n \geqslant 0$ and domain $U \subset \mathbb{F}$, that is,

$$\mathbb{P}_n(U) = \big\{ \boldsymbol{p}(x) = a_0 + a_1 x + \cdots + a_n x^n, \text{ with } a_0, \cdots, a_n \in \mathbb{F} \text{ and } x \in U \subset \mathbb{F} \big\}.$$

The set $\mathbb{P}_n(U)$ together with the addition of polynomials $(\boldsymbol{p} + \boldsymbol{q})(x) = \boldsymbol{p}(x) + \boldsymbol{q}(x)$ and the scalar multiplication $(a\boldsymbol{p})(x) = a\,\boldsymbol{p}(x)$ is a vector space. In the case $U = \mathbb{R}$ we use the notation $\mathbb{P}_n = \mathbb{P}_n(\mathbb{R})$. ◁

EXAMPLE 4.1.4: Let $C^k\big([a,b], \mathbb{F}\big)$ be the set of scalar valued functions with domain $[a,b] \subset \mathbb{R}$ with the k-th derivative being a continuous function, that is,

$$C^k\big([a,b], \mathbb{F}\big) = \big\{ \boldsymbol{f} : [a,b] \to \mathbb{F} \quad \text{such that} \quad \boldsymbol{f}^{(k)} \quad \text{is continuous} \big\}.$$

The set $C^k\big([a,b], \mathbb{F}\big)$ together with the addition of functions $(\boldsymbol{f} + \boldsymbol{g})(x) = \boldsymbol{f}(x) + \boldsymbol{g}(x)$ and the scalar multiplication $(a\boldsymbol{f})(x) = a\,\boldsymbol{f}(x)$ is a vector space. The particular case $C^k(\mathbb{R}, \mathbb{R})$ is denoted simply as C^k. The set of real valued continuous function is then C^0. ◁

EXAMPLE 4.1.5: Let ℓ be the set of absolute convergent series, that is,

$$\ell = \Big\{ \boldsymbol{a} = \sum a_n \;:\; a_n \in \mathbb{F} \text{ and } \sum_{n=0}^{\infty} |a_n| \text{ exists} \Big\}.$$

The set ℓ with the addition of series $\boldsymbol{a} + \boldsymbol{b} = \sum (a_n + b_n)$ and the scalar multiplication $c\boldsymbol{a} = \sum c\, a_n$ is a vector space. ◁

The properties (A1)-(M5) given in the definition of vector space are not redundant. As an example, these properties do not include the condition that the neutral element $\boldsymbol{0}$ is unique, since it follows from the definition.

Theorem 4.1.2. *The element $\boldsymbol{0}$ in a vector space is unique.*

Proof Theorem 4.1.2: Suppose that there exist two neutral elements $\boldsymbol{0}_1$ and $\boldsymbol{0}_2$ in the vector space V, that is,

$$\boldsymbol{0}_1 + \boldsymbol{u} = \boldsymbol{u} \quad \text{and} \quad \boldsymbol{0}_2 + \boldsymbol{u} = \boldsymbol{u} \quad \text{for all} \quad \boldsymbol{u} \in V$$

Taking $\boldsymbol{u} = \boldsymbol{0}_2$ in the first equation above, and $\boldsymbol{u} = \boldsymbol{0}_1$ in the second equation above we obtain that

$$\boldsymbol{0}_1 + \boldsymbol{0}_2 = \boldsymbol{0}_2, \qquad \boldsymbol{0}_2 + \boldsymbol{0}_1 = \boldsymbol{0}_1.$$

These equations above simply that the two neutral elements must be the same, since

$$\boldsymbol{0}_2 = \boldsymbol{0}_1 + \boldsymbol{0}_2 = \boldsymbol{0}_2 + \boldsymbol{0}_1 = \boldsymbol{0}_1;$$

where in the second equation we used that the addition operation is commutative. This establishes the Theorem. □

Theorem 4.1.3. *It holds that $0\,\boldsymbol{u} = \boldsymbol{0}$ for all element $\boldsymbol{u}$ in a vector space V.*

Proof Theorem 4.1.3: For every $\boldsymbol{u} \in V$ holds

$$\boldsymbol{u} = 1\,\boldsymbol{u} = (1+0)\boldsymbol{u} = 1\,\boldsymbol{u} + 0\,\boldsymbol{u} = \boldsymbol{u} + 0\,\boldsymbol{u} = 0\,\boldsymbol{u} + \boldsymbol{u} \quad \Rightarrow \quad \boldsymbol{u} = 0\,\boldsymbol{u} + \boldsymbol{u}.$$

This last equation says that $0\boldsymbol{u}$ is a neutral element, $\boldsymbol{0}$. Theorem 4.1.2 says that the neutral element is unique, so we conclude that, for all $\boldsymbol{u} \in V$ holds that

$$0\,\boldsymbol{u} = \boldsymbol{0}.$$

This establishes the Theorem. □

Also notice that the property (A5) in the definition of vector space says that the opposite element exists, but it does not say whether it is unique. The opposite element is actually unique.

Theorem 4.1.4. *The opposite element $-\boldsymbol{u}$ in a vector space is unique.*

Proof Theorem 4.1.4: Suppose there are two opposite elements $-\boldsymbol{u}_1$ and $-\boldsymbol{u}_2$ to the element $\boldsymbol{u} \in V$, that is,

$$\boldsymbol{u} + (-\boldsymbol{u}_1) = \boldsymbol{0}, \qquad \boldsymbol{u} + (-\boldsymbol{u}_2) = \boldsymbol{0}.$$

Therefore,

$$\begin{aligned}(-\boldsymbol{u}_1) &= \boldsymbol{0} + (-\boldsymbol{u}_1) \\ &= u + (-\boldsymbol{u}_2) + (-\boldsymbol{u}_1) \\ &= (-\boldsymbol{u}_2) + \boldsymbol{u} + (-\boldsymbol{u}_1) \\ &= (-\boldsymbol{u}_2) + \boldsymbol{0} \\ &= \boldsymbol{0} + (-\boldsymbol{u}_2) \\ &= (-\boldsymbol{u}_2) \quad \Rightarrow \quad (-\boldsymbol{u}_1) = (-\boldsymbol{u}_2).\end{aligned}$$

This establishes the Theorem. □

Finally, notice that the element $(-\boldsymbol{u})$ opposite to $\boldsymbol{u}$ is actually the element $(-1)\boldsymbol{u}$.

Theorem 4.1.5. *It holds that $(-1)\,\boldsymbol{u} = (-\boldsymbol{u})$.*

Proof Theorem 4.1.5:

$$\boldsymbol{0} = 0\,\boldsymbol{u} = (1-1)\,\boldsymbol{u} = 1\,\boldsymbol{u} + (-1)\,\boldsymbol{u} = \boldsymbol{u} + (-1)\,\boldsymbol{u}.$$

Hence $(-1)\,\boldsymbol{u}$ is an opposite element of $\boldsymbol{u}$. Since Theorem 4.1.4 says that the opposite element is unique, we conclude that $(-1)\,\boldsymbol{u} = (-\boldsymbol{u})$. This establishes the Theorem. □

4.1.1. **Subspaces.** We now introduce the notion of a subspace of a vector space, which is essentially a smaller vector space inside the original vector space. Subspaces are important in physics, since physical processes frequently take place not inside the whole space but in a particular subspace. For instance, planetary motion does not take plane in the whole space but it is confined to a plane.

Definition 4.1.6. *The subset $W \subset V$ of a vector space V over $\mathbb{F}$ is called a **subspace** of V iff for all $\boldsymbol{u}, \boldsymbol{v} \in W$ and all $a, b \in \mathbb{F}$ holds that $a\boldsymbol{u} + b\boldsymbol{v} \in W$.*

A subspace is a particular type of set in a vector space. Is a set where all possible linear combinations of two elements in the set results in another element in the same set. In other words, elements outside the set cannot be reached by linear combinations of elements within the set. For this reason a subspace is called a *closed set under linear combinations*. The following statement is usually helpful

Theorem 4.1.7. *If $W \subset V$ is a subspace of a vector space V, then $\boldsymbol{0} \in W$.*

This statement says that $\boldsymbol{0} \notin W$ implies that W *is not a subspace*. However, if actually $\boldsymbol{0} \in W$, this fact alone does not prove that W is a subspace. One must show that W is closed under linear combinations.

Proof of Theorem 4.1.7: Since W is closed under linear combinations, given any element $\boldsymbol{u} \in W$, the trivial linear combination $0\,\boldsymbol{u} = \boldsymbol{0}$ must belong to W, hence $\boldsymbol{0} \in W$. This establishes the Theorem. □

EXAMPLE 4.1.6: Show that the set $W \subset \mathbb{R}^3$ given by $W = \left\{ \mathsf{u} = [u_i] \in \mathbb{R}^3 \ : \ u_3 = 0 \right\}$ is a subspace of $\mathbb{R}^3$:

SOLUTION: Given two arbitrary elements $\mathsf{u}, \mathsf{v} \in W$ we must show that $a\mathsf{u} + b\mathsf{v} \in W$ for all $a, b \in \mathbb{R}$. Since $\mathsf{u}, \mathsf{v} \in W$ we know that

$$\mathsf{u} = \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}.$$

Therefore

$$a\mathsf{u} + b\mathsf{v} = \begin{bmatrix} au_1 + bv_1 \\ au_2 + bv_2 \\ 0 \end{bmatrix} \in W,$$

since the third component vanishes, which makes the linear combination an element in W. Hence, W is a subspace of $\mathbb{R}^3$. In Fig. 29 we see the plane $u_3 = 0$. It is a subspace, since not only $\mathsf{0} \in W$, but any linear combination of vectors on the plane stays on the plane. ◁

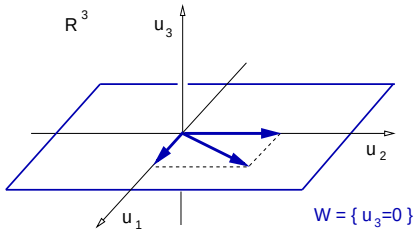

FIGURE 29. The horizontal plane $u_3 = 0$ is a subspace of $\mathbb{R}^3$.

EXAMPLE 4.1.7: Show that the set $W = \{\mathsf{u} = [u_i] \in \mathbb{R}^2 \ : \ u_2 = 1\}$ is not a subspace of $\mathbb{R}^2$.

SOLUTION: The set W is not a subspace, since $\mathsf{0} \notin W$. This is enough to show that W is not a subspace. Another proof is that the addition of two vectors in the set is a vector outside the set, as can be seen by the following calculation,

$$\mathsf{u} = \begin{bmatrix} u_1 \\ 1 \end{bmatrix} \in W, \quad \mathsf{v} = \begin{bmatrix} v_1 \\ 1 \end{bmatrix} \in W \quad \Rightarrow \quad \mathsf{u} + \mathsf{v} = \begin{bmatrix} u_1 + v_1 \\ 2 \end{bmatrix} \notin W.$$

The second component in the addition above is 2, not 1, hence this addition does not belong to W. An example of this calculation is given in Fig. 30. ◁

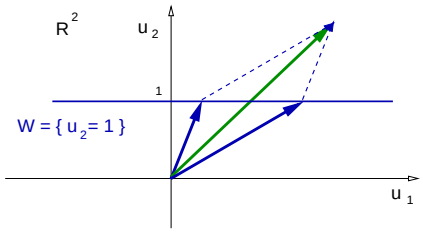

FIGURE 30. The horizontal line $u_2 = 1$ is not a subspace of $\mathbb{R}^2$.

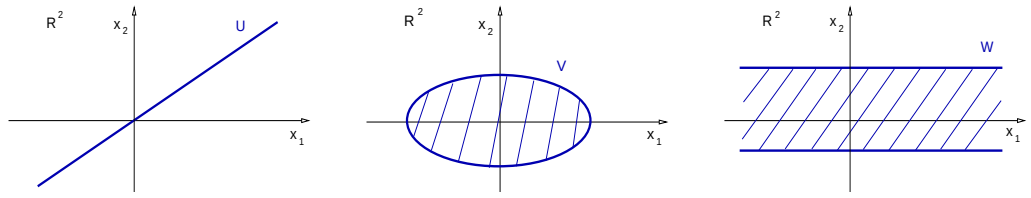

FIGURE 31. Three subsets, U, V, and W, of $\mathbb{R}^2$. Only the set U is a subspace.

EXAMPLE 4.1.8: Determine which one of the sets given in Fig. 31 is a subspace of $\mathbb{R}^2$.

SOLUTION: The set U is a vector space, since any linear combination of vectors parallel to the line is again a vector parallel to the line. The sets V and W are not subspaces, since given a vector $\boldsymbol{u}$ in these spaces, a the vector $a\boldsymbol{u}$ does not belong to these sets for a number $a \in \mathbb{R}$ big enough. This argument is sketched in Fig. 32.

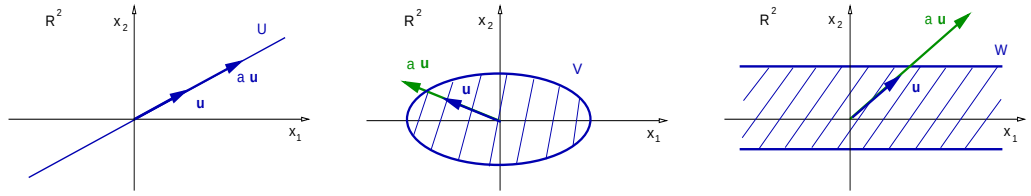

FIGURE 32. Three subsets, U, V, and W, of $\mathbb{R}^2$. Only the set U is a subspace.

◁

4.1.2. **The span of finite sets.** If a set is not a subspace there is a way to increase it into a subspace. Define a new set including all possible linear combinations of elements in the old set.

Definition 4.1.8. *The **span** of a finite set $S = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ in a vector space V over $\mathbb{F}$, denoted as $\text{Span}(S)$, is the set given by*

$$\text{Span}(S) = \{\boldsymbol{u} \in V \ : \ \boldsymbol{u} = c_1\boldsymbol{u}_1 + \cdots + c_n\boldsymbol{u}_n, \quad c_1, \cdots, c_n \in \mathbb{F}\}.$$

The following result remarks that the span of a set is a subspace.

Theorem 4.1.9. *Given a finite set S in a vector space V, $\text{Span}(S)$ is a subspace of V.*

Proof of Theorem 4.1.9: Since $\text{Span}(S)$ contains all possible linear combinations of the elements in S, then $\text{Span}(S)$ is closed under linear combinations. This establishes the Theorem. □

EXAMPLE 4.1.9: The subspace $\text{Span}(\{\boldsymbol{v}\})$, that is, the set of all possible linear combinations of the vector $\boldsymbol{v}$, is formed by all vectors of the form $c\boldsymbol{v}$, and these vectors belong to a line containing $\boldsymbol{v}$. The subspace $\text{Span}(\{\boldsymbol{v}, \boldsymbol{w}\})$, that is, the set of all linear combinations of two vectors $\boldsymbol{v}$, $\boldsymbol{w}$, is the plane containing both vectors $\boldsymbol{v}$ and $\boldsymbol{w}$. See Fig. 33 for the case of the vector spaces $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. ◁

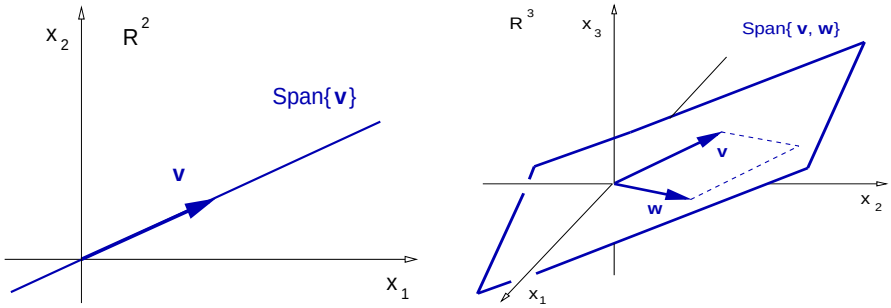

FIGURE 33. Examples of the span of a set of a single vector, and the span of a linearly independent set of two vectors.

4.1.3. **Algebra of subspaces.** We now show that the intersection of two subspaces is again a subspace. However, the union of two subspaces is not, in general, a subspace. The smaller subspace containing the union of two subspaces is precisely the span of the union. We then define the addition of two subspaces as the span of the union of two subspaces.

Theorem 4.1.10. *If W_1 and W_2 are subspaces of a vector space V, then $W_1 \cap W_2 \subset V$ is also a subspace of V.*

Proof of Theorem 4.1.10: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be any two elements in $W_1 \cap W_2$. This means that $\boldsymbol{u}$, $\boldsymbol{v} \in W_1$, which is a subspace, so any linear combination $(a\boldsymbol{u} + b\boldsymbol{v}) \in W_1$. Since $\boldsymbol{u}$, $\boldsymbol{v}$ belong to $W_1 \cap W_2$ they also belong to W_2, which is a subspace, so any linear combination $(a\boldsymbol{u} + b\boldsymbol{v}) \in W_2$. Therefore, any linear combination $(a\boldsymbol{u} + b\boldsymbol{v}) \in W_1 \cap W_2$. This establishes the Theorem. □

EXAMPLE 4.1.10: The sketch in Fig. 34 shows the intersection of two subspaces in $\mathbb{R}^3$, a plane and a line. In this case the intersection is the former line, so the intersection is a subspace.

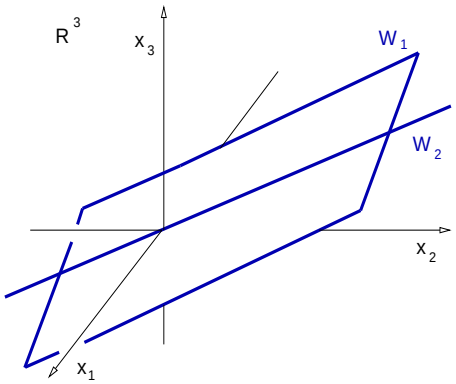

FIGURE 34. Intersection of two subspaces, W_1 and W_2 in $\mathbb{R}^3$. Since the line W_2 is included into the plane W_1, we have that $W_1 \cap W_2 = W_2$.

◁

While the intersection of two subspaces is always a subspace, their union is, in general, not a subspace, unless one subspace is contained into the other.

EXAMPLE 4.1.11: Consider the vector space $V = \mathbb{R}^2$, and the subspaces W_1 and W_2 given by the lines sketched in Fig. 35. Their union is the set formed by these two lines. This set is not a subspace, since the addition of the vectors $\boldsymbol{u}_1 \in W_1$ with $\boldsymbol{u}_2 \in W_2$ does not belong to $W_1 \cup W_2$, as is it shown in Fig. 35.

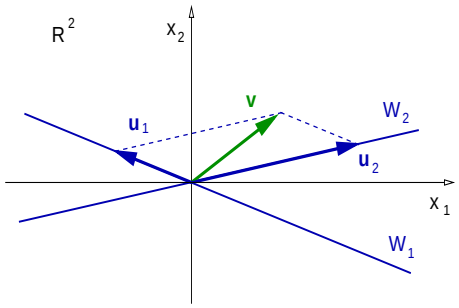

FIGURE 35. The union of the subspaces W_1 and W_2 is the set formed by these two lines. This is not a subspace, since the addition of $\boldsymbol{u}_1 \in W_1$ and $\boldsymbol{u}_2 \in W_2$ is the vector $\boldsymbol{v}$ which does not belong to $W_1 \cup W_2$.

◁

Although the union of two subspaces is not always a subspace, it is possible to enlarge the union into a subspace. The idea is to incorporate all possible additions of vectors in the two original subspaces, and the result is called the addition of the two subspaces. Here is the precise definition.

Definition 4.1.11. *The **addition of the subspaces** W_1, W_2 in a vector space V, denoted as $W_1 + W_2$, is the set given by*

$$W_1 + W_2 = \big\{ \boldsymbol{u} \in V \ : \ \boldsymbol{u} = \boldsymbol{w}_1 + \boldsymbol{w}_2 \text{ with } \boldsymbol{w}_1 \in W_1, \ \boldsymbol{w}_2 \in W_2 \big\}.$$

The following result remarks that the addition of subspaces is again a subspace.

Theorem 4.1.12. *If W_1 and W_2 are subspaces of a vector space V, then the addition $W_1 + W_2$ is also a subspace of V.*

Proof of Theorem 4.1.12: Suppose that $\boldsymbol{x} \in W_1 + W_2$ and $\boldsymbol{y} \in W_1 + W_2$. We must show that any linear combination $a\boldsymbol{x} + b\boldsymbol{y}$ also belongs to $W_1 + W_2$. This is the case, by the following argument. Since $\boldsymbol{x} \in W_1 + W_2$, there exist $\boldsymbol{x}_1 \in W_1$ and $\boldsymbol{x}_2 \in W_2$ such that $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$. Analogously, since $\boldsymbol{y} \in W_1 + W_2$, there exist $\boldsymbol{y}_1 \in W_1$ and $\boldsymbol{y}_2 \in W_2$ such that $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$. Now any linear combination of $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfies

$$\begin{aligned} a\boldsymbol{x} + b\boldsymbol{y} &= a(\boldsymbol{x}_1 + \boldsymbol{x}_2) + b(\boldsymbol{y}_1 + \boldsymbol{y}_2) \\ &= (a\boldsymbol{x}_1 + b\boldsymbol{y}_1) + (a\boldsymbol{x}_2 + b\boldsymbol{y}_2) \end{aligned}$$

Since W_1 and W_2 are subspaces, $(a\boldsymbol{x}_1 + b\boldsymbol{y}_2) \in W_1$, and $(a\boldsymbol{x}_2 + b\boldsymbol{y}_2) \in W_2$. Therefore, the equation above says that $(a\boldsymbol{x} + b\boldsymbol{y}) \in W_1 + W_2$. This establishes the Theorem. □

EXAMPLE 4.1.12: The sketch in Fig. 36 shows the union and the addition of two subspaces in $\mathbb{R}^3$, each subspace given by a line through the origin. While the union is not a subspace, their addition is the plane containing both lines, which is a subspace. Given any non-zero vector $\boldsymbol{w}_1 \in W_1$ and any other non-zero vector $\boldsymbol{w}_2 \in W_2$, one can verify that the sum of two subspaces is the span of $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$, that is,

$$W_1 + W_2 = \text{Span}\big(\{\boldsymbol{w}_1\} \cup \{\boldsymbol{w}_2\}\big).$$

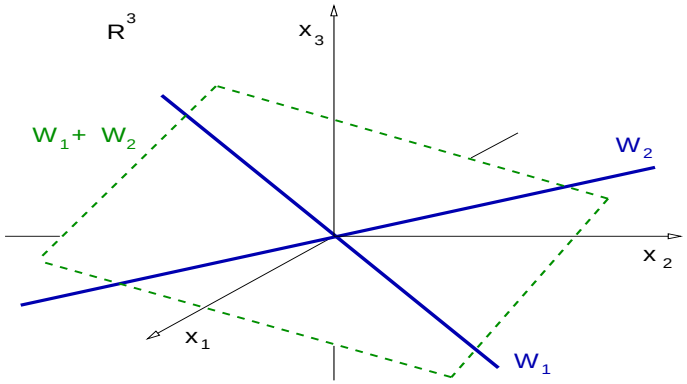

FIGURE 36. Union and addition of the subspaces W_1 and W_2 in $\mathbb{R}^3$. The union is not a subspace, while the addition is a subspace of $\mathbb{R}^3$.

4.1.4. **Internal direct sums.** This is a particular case of the addition of subspaces. It is called internal direct sum in order to differentiate it from another type of direct sum found in the literature. The latter, also called external direct sum, is a sum of different vector spaces, and it is a way to construct new vector spaces from old ones. We do not discuss this type of direct sums here. From now on, direct sum in these notes means the internal direct sum of subspaces inside a vector space.

Definition 4.1.13. *Given a vector space V, we say that V is the **internal direct sum** of the subspaces W_1, $W_2 \subset V$, denoted as $V = W_1 \oplus W_2$, iff every vector $\boldsymbol{v} \in V$ can be written in a unique way, except for order, as a sum of vectors from W_1 and W_2.*

A crucial part in the definition above is the uniqueness of the decomposition of every vector $\boldsymbol{v} \in V$ as a sum of a vector in W_1 plus a vector in W_2. By uniqueness we mean the following: For every $\boldsymbol{v} \in V$ exist $\boldsymbol{w}_1 \in \mathbb{W}_1$ and $\boldsymbol{w}_2 \in W_2$ such that $\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2$, and if $\boldsymbol{v} = \tilde{\boldsymbol{w}}_1 + \tilde{\boldsymbol{w}}_2$ with $\tilde{\boldsymbol{w}}_1 \in W_1$ and $\tilde{\boldsymbol{w}}_2 \in W_2$, then $\boldsymbol{w}_1 = \tilde{\boldsymbol{w}}_1$ and $\boldsymbol{w}_2 = \tilde{\boldsymbol{w}}_2$. In the case that $V = W_1 \oplus W_2$ we say that W_1 and W_2 are *direct summands* of V, and we also say that W_1 is the *direct complement* of W_2 in V. There is an useful characterization of internal direct sums.

Theorem 4.1.14. *A vector space V is the direct sum of subspaces W_1 and W_2 iff holds both $V = W_1 + W_2$ and $W_1 \cap W_2 = \{\boldsymbol{0}\}$.*

Proof of Theorem 4.1.14:
($\Rightarrow$) If $V = W_1 \oplus W_2$, then it implies that $V = W_1 + W_2$. Suppose that $\boldsymbol{v} \in W_1 \cap W_2$, then on the one hand, there exists $\boldsymbol{w}_1 \in W_1$ such that $\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{0}$; on the other hand, there is $\boldsymbol{w}_2 \in W_2$ such that $\boldsymbol{v} = \boldsymbol{0} + \boldsymbol{w}_2$. Therefore, $\boldsymbol{w}_1 = \boldsymbol{0}$ and $\boldsymbol{w}_2 = \boldsymbol{0}$, so $W_1 \cap W_2 = \{\boldsymbol{0}\}$.
($\Leftarrow$) Since $V = W_1 + W_2$, for every $\boldsymbol{v} \in V$ there exist $\boldsymbol{w}_1 \in W_1$ and $\boldsymbol{w}_2 \in W_2$ such that $\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2$. Suppose there exists other vectors $\tilde{\boldsymbol{w}}_1 \in W_1$ and $\tilde{\boldsymbol{w}}_2 \in W_2$ such that $\boldsymbol{v} = \tilde{\boldsymbol{w}}_1 + \tilde{\boldsymbol{w}}_2$. Then,

$$\boldsymbol{0} = (\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) + (\boldsymbol{w}_2 - \tilde{\boldsymbol{w}}_2) \quad \Leftrightarrow \quad (\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) = -(\boldsymbol{w}_2 - \tilde{\boldsymbol{w}}_2),$$

Therefore $(\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) \in W_2$ and so $(\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) \in W_1 \cap W_2$. Since $W_1 \cap W_2 = \{\boldsymbol{0}\}$, we then conclude that $\boldsymbol{w}_1 = \tilde{\boldsymbol{w}}_1$, which also says $\boldsymbol{w}_2 = \tilde{\boldsymbol{w}}_2$. Then $V = W_1 \oplus W_2$. This establishes the Theorem. □

EXAMPLE 4.1.13: Denote by Sym and SkewSym the sets of all symmetric and all skew-symmetric $n \times n$ matrices. Show that $\mathbb{F}^{n,n} = \text{Sym} \oplus \text{SkewSym}$.

SOLUTION: Given any matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, holds

$$\mathsf{A} = \mathsf{A} + \frac{1}{2}\big(\mathsf{A}^T - \mathsf{A}^T\big) = \frac{1}{2}\big(\mathsf{A} + \mathsf{A}^T\big) + \frac{1}{2}\big(\mathsf{A} - \mathsf{A}^T\big).$$

We then can decompose matrix as $\mathsf{A} = \mathsf{B} + \mathsf{C}$, where matrix $\mathsf{B} = \big(\mathsf{A} + \mathsf{A}^T\big)/2 \in \text{Sym}$ while matrix $\mathsf{C} = \big(\mathsf{A} - \mathsf{A}^T\big)/2 \in \text{SkewSym}$. That is, we can write any square matrix as a symmetric matrix plus a skew-symmetric matrix, hence $\mathbb{F}^{n,n} \subset \text{Sym} + \text{SkewSym}$. The other inclusion is obvious, that is, $\text{Sym} + \text{SkewSym} \subset \mathbb{F}^{n,n}$, because each term in the sum is a subset of $\mathbb{F}^{n,n}$. So, we conclude that

$$\mathbb{F}^{n,n} = \text{Sym} + \text{SkewSym}.$$

Now we must show that $\text{Sym} \cap \text{SkewSym} = \{\boldsymbol{0}\}$. This is the case, as the following argument shows. If matrix $\mathsf{D} \in \text{Sym} \cap \text{SkewSym}$, then matrix D is symmetric, $\mathsf{D} = \mathsf{D}^T$, but matrix D is also skew-symmetric, $\mathsf{D} = -\mathsf{D}^T$. This implies that $\mathsf{D} = -\mathsf{D}$, that is, $\mathsf{D} = \mathsf{0}$, proving our assertion that $\text{Sym} \cap \text{SkewSym} = \{\boldsymbol{0}\}$. We then conclude that

$$\mathbb{F}^{n,n} = \text{Sym} \oplus \text{SkewSym}.$$

◁

4.1.5. **Exercises.**

4.1.1.- Determine which of the following subsets of $\mathbb{R}^n$, with $n \geqslant 1$, are in fact subspaces. Justify your answers.

(a) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_i \geqslant 0 \quad i = 1, \cdots, n\}$;
(b) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 = 0\}$;
(c) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 x_2 = 0 \quad n \geqslant 2\}$;
(d) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 + \cdots + x_n = 0\}$;
(e) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 + \cdots + x_n = 1\}$;
(f) $\{\mathsf{x} \in \mathbb{R}^n \ : \ \mathsf{A}\mathsf{x} = \mathsf{b}, \ \mathsf{A} \neq \mathsf{0}, \ \mathsf{b} \neq \mathsf{0}\}$.

4.1.2.- Determine which of the following subsets of $\mathbb{F}^{n,n}$, with $n \geqslant 1$, are in fact subspaces. Justify your answers.

(a) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A} = \mathsf{A}^T\}$;
(b) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}$ invertible$\}$;
(c) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}$ not invertible$\}$;
(d) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}$ upper-triangular$\}$;
(e) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}^2 = \mathsf{A}\}$;
(f) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathrm{tr}\,(\mathsf{A}) = 0\}$.
(g) Given a matrix $\mathsf{X} \in \mathbb{F}^{n,n}$, define $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ [\mathsf{A}, \mathsf{X}] = \mathsf{0}\}$.

4.1.3.- Find $W_1 + W_2 \subset \mathbb{R}^3$, where W_1 is a plane passing through the origin in $\mathbb{R}^3$ and W_2 is a line passing through the origin in $\mathbb{R}^3$ not contained in W_1.

4.1.4.- Sketch a picture of the subspaces spanned by the following vectors:

(a) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ -6 \end{bmatrix} \right\}$;

(b) $\left\{ \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\}$;

(c) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

4.1.5.- Given two finite subsets S_1, S_2 in a vector space V, show that

$$\mathrm{Span}(S_1 \cup S_2) = \mathrm{Span}(S_1) + \mathrm{Span}(S_2).$$

4.1.6.- Let $W_1 \subset \mathbb{R}^3$ be the subspace

$$W_1 = \mathrm{Span}\Big(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\} \Big).$$

Find a subspace $W_2 \subset \mathbb{R}^3$ such that $\mathbb{R}^3 = W_1 \oplus W_2$.

4.2. Linear dependence

4.2.1. **Linearly dependent sets.** In this Section we present the notion of a linearly dependent set of vectors. If one of the vectors in the set is a linear combination of the other vectors in the set, then the set is called linearly dependent. If this is not the case, the set is called linearly independent. This notion plays a crucial role in Sect. 4.3 to define a basis of a vector space. Bases are very useful in part because every vector in the vector space can be decomposed in a unique way as a linear combination of the basis elements. Bases also provide a precise way to measure the size of a vector space.

Definition 4.2.1. *A finite set of vectors $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ in a vector space is called* ***linearly dependent*** *iff there exists a set of scalars $\{c_1, \cdots, c_k\}$, not all of them zero, such that,*

$$c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k = \boldsymbol{0}. \tag{4.1}$$

On the other hand, the set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ is called ***linearly independent*** *iff Eq. (4.1) implies that every scalar vanishes, $c_1 = \cdots = c_k = 0$.*

The wording in this definition is carefully chosen to cover the case of the empty set. The result is that the empty set is linearly independent. It might seems strange, but this result fits well with the rest of the theory. On the other hand, the set $\{\boldsymbol{0}\}$ is linearly dependent, since $c_1 \boldsymbol{0} = \boldsymbol{0}$ for any $c_1 \neq 0$. Moreover, any set containing the zero vector is also linearly dependent.

Linear dependence or independence are properties of a set of vectors. There is no meaning to a vector to be linearly dependent, or independent. And there is no meaning of a set of linearly dependent vectors, as well as a set of linearly independent vectors. What is meaningful is to talk of a linearly dependent or independent set of vectors.

Example 4.2.1: Show that the set $S \subset \mathbb{R}^2$ below is linearly dependent,

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}.$$

Solution: It is clear that

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since $c_1 = 2$, $c_2 = 3$, and $c_3 = -1$ are non-zero, the set S is linearly dependent. ◁

It will be convenient to have the concept of a linearly dependent or independent set containing infinitely many vectors.

Definition 4.2.2. *An infinite set of vectors $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\}$ in a vector space V is called* ***linearly independent*** *iff every finite subset of S is linearly independent. Otherwise, the infinite set S is called linearly dependent.*

Example 4.2.2: Consider the vector space $V = C^\infty\big([-\ell, \ell], \mathbb{R}\big)$, that is, the space of infinitely differentiable real-valued functions defined on the domain $[-\ell, \ell] \subset \mathbb{R}$ with the usual operation of linear combination of functions. This vector space contains linearly independent sets with infinitely many vectors. One example is the infinite sets S_1 below, which is linearly independent,

$$S_1 = \{1, x, x^2, \cdots, x^n, \cdots\}.$$

Another example is the infinite set S_2, which is also linearly independent,

$$S_2 = \left\{1, \cos\Big(\frac{n\pi x}{\ell}\Big), \sin\Big(\frac{n\pi x}{\ell}\Big)\right\}_{n=1}^{\infty}.$$

◁

4.2.2. **Main properties.** As we have seen in the Example 4.2.1 above, in a linearly dependent set there is always at least one vector that is a linear combination of the other vectors in the set. This is simple to see from the Definition 4.2.1. Since not all the coefficients c_i are zero in a linearly dependent set, let us suppose that $c_j \neq 0$; then from the Eq. (4.1) we obtain

$$\boldsymbol{v}_j = -\frac{1}{c_j}\big[c_1\boldsymbol{v}_1 + \cdots + c_{j-1}\boldsymbol{v}_{j-1} + c_{j+1}\boldsymbol{v}_{j+1} + \cdots + c_k\boldsymbol{v}_k\big],$$

that is, $\boldsymbol{v}_j$ is a linear combination of the other vectors in the set.

Theorem 4.2.3. *The set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ is linearly dependent with the vector $\boldsymbol{v}_k$ being a linear combination of the the remaining $k-1$ vectors iff*

$$\text{Span}\big(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}\big) = \text{Span}\big(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}\}\big).$$

This Theorem captures the idea behind the notion of a linearly dependent set: A finite set is linearly dependent iff there exists a smaller set with the same span. In this sense the vector $\boldsymbol{v}_k$ in the Proposition above is redundant with respect to linear combinations.

Proof of Theorem 4.2.3: Let $S_k = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ and $S_{k-1} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}\}$.

On the one hand, if $\boldsymbol{v}_k$ is a linear combination of the other vectors in S, then for every $\boldsymbol{x} \in \text{Span}(S_k)$ can be expressed as an element in $\text{Span}(S_{k-1})$ simply by replacing $\boldsymbol{v}_k$ in terms of the vectors in $\tilde{S}$. This shows that $\text{Span}(S_k) \subset \text{Span}(S_{k-1})$. The other inclusion is trivial, so $\text{Span}(S_k) = \text{Span}(S_{k-1})$.

On the other hand, if $\text{Span}(S_k) = \text{Span}(S_{k-1})$, this means that $\boldsymbol{v}_k$ is a linear combination of the elements in S_{k-1}. Therefore, the set S_k is linearly dependent. This establishes the Theorem. □

EXAMPLE 4.2.3: Consider the set $S \subset \mathbb{R}^3$ given by $S = \left\{ \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix}, \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \\ -3 \end{bmatrix} \right\}$. Find a set $\tilde{S} \subset S$ having the smallest number of vectors such that $\text{Span}(\tilde{S}) = \text{Span}(S)$.

SOLUTION: We have to find all the redundant vectors in S with respect to linear combinations. In other words, with have to find a linearly independent subset of $\tilde{S} \subset S$ such that $\text{Span}(\tilde{S}) = \text{Span}(S)$. The calculation we must do is to find the non-zero coefficients c_i in the solution of the equation

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} + c_3 \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix} + c_4 \begin{bmatrix} -4 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}.$$

Hence, we must find the reduced echelon form of the coefficient matrix above, that is,

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 & 2 \\ 0 & -2 & -5 & -3 \\ 0 & 2 & 5 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 & 2 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & \frac{5}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathsf{E}_\mathsf{A}.$$

This means that the solution for the coefficients is

$$c_1 = -6c_3 - 5c_4, \qquad c_2 = -\frac{5}{2}c_3 - \frac{3}{2}c_4, \qquad c_3, c_4 \text{ free variables.}$$

Choosing:

$$c_4 = 0,\ c_3 = 2 \quad \Rightarrow \quad c_1 = -12,\ c_2 = -5 \quad \Rightarrow \quad -12 \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} - 5 \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} + 2 \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$c_4 = 2,\ c_3 = 0 \quad \Rightarrow \quad c_1 = -10,\ c_2 = -3 \quad \Rightarrow \quad -10\begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} - 3\begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} + 2\begin{bmatrix} -4 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

We can interpret this result thinking that the third and fourth vectors in matrix A are linear combination of the first two vectors. Therefore, a linearly independent subset of S having its same span is given by

$$\tilde{S} = \left\{ \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} \right\}.$$

Notice that all the information to find $\tilde{S}$ is in matrix $\mathsf{E_A}$, the reduced echelon form of matrix A,

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & \frac{5}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathsf{E_A}.$$

The columns with pivots in $\mathsf{E_A}$ determine the column vectors in A that form a linearly independent set. The non-pivot columns in $\mathsf{E_A}$ determine the column vectors in A that are linear combination of the vectors in the linearly independent set. The factors of these linear combinations are precisely the component of the non-pivot vectors in $\mathsf{E_A}$. For example, the last column vector in $\mathsf{E_A}$ has components 5 and 3/2, and these are precisely the coefficients in the linear combination:

$$\begin{bmatrix} -4 \\ 1 \\ -3 \end{bmatrix} = 5\begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix}.$$

◁

In Example 4.2.3 we answered a question about the linear independence of a set $S = \{\mathsf{v}_1, \cdots, \mathsf{v}_n\} \subset \mathbb{F}^n$ by studying the properties of a matrix having these vectors a column vectors, that is, $\mathsf{A} = [\mathsf{v}_1, \cdots, \mathsf{v}_n]$. It turns out that this is a good idea and the following result summarizes few useful relations.

Theorem 4.2.4. *Given $\mathsf{A} = [\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}] \in \mathbb{F}^{m,n}$, the following statements are equivalent:*
(a) The column vectors set $\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\} \subset \mathbb{F}^m$ is linearly independent;
(b) $N(\mathsf{A}) = \{\mathbf{0}\}$;
(c) $\text{rank}(\mathsf{A}) = n$
In the case $\mathsf{A} \in \mathbb{F}^{n,n}$, the set $\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\} \subset \mathbb{F}^n$ is linearly independent iff A is invertible.

Proof of Theorem 4.2.4: Let us denote by $S = \{\mathsf{v}_1, \cdots, \mathsf{v}_n\} \subset \mathbb{F}^m$ a set of vectors in a vector space, and introduce the matrix $\mathsf{A} = [\mathsf{v}_1, \cdots, \mathsf{v}_n]$. The set S is linearly independent iff only solution $\mathsf{c} \in \mathbb{R}^n$ to the equation $\mathsf{Ac} = \mathsf{0}$ is the trivial solution $\mathsf{c} = \mathsf{0}$. This is equivalent to say that $N(\mathsf{A}) = \{\mathsf{0}\}$. This is equivalent to say that $\mathsf{E_A}$ has n pivot columns, which is equivalent to say that $\text{rank}(\mathsf{A}) = n$. The furthermore part is straightforward, since an $n \times n$ matrix A is invertible iff $\text{rank}(\mathsf{A}) = n$. This establishes the Theorem. □

Further reading. See Section 4.3 in Meyer's book [3].

4.2.3. **Exercises.**

4.2.1.- Determine which of the following sets is linearly independent. For those who are linearly dependent, express one vector as a linear combination of the other vectors in the set.

(a) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix} \right\}$;

(b) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$;

(c) $\left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\}$.

4.2.2.- Let $\mathsf{A} = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 2 & 1 & 2 \\ 6 & 3 & 2 & 3 \end{bmatrix}$.

(a) Find a linearly independent set containing the largest possible number of columns of A.

(b) Find how many linearly independent sets can be constructed using any number of column vectors of A.

4.2.3.- Show that any set containing the zero vector must be linearly dependent.

4.2.4.- Given a vector space V, prove the following: If the set

$$\{\boldsymbol{v},\ \boldsymbol{w}\} \subset V$$

is linearly independent, then so is

$$\big\{(\boldsymbol{v}+\boldsymbol{w}),\ (\boldsymbol{v}-\boldsymbol{w})\big\}.$$

4.2.5.- Determine whether the set

$$\left\{ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix} \right\} \subset \mathbb{R}^{2,2}$$

is linearly independent of dependent.

4.2.6.- Show that the following set in $\mathbb{P}_2$ is linearly dependent,

$$\{1,\ x,\ x^2,\ 1+x+x^2\}.$$

4.2.7.- Determine whether $S \subset \mathbb{P}_2$ is a linearly independent set, where

$$S = \big\{1+x+x^2, 2x-3x^2, 2+x\big\}.$$

4.3. Bases and dimension

In this Section we introduce a notion that quantifies the size of a vector space. Before doing that, however, we need to separate two main cases, the vector spaces we call finite dimensional from those called infinite dimensional. In the first case, finite dimensional vector spaces, we introduce the notion of a basis. This is a particular type of set in the vector space that is small enough to be a linearly independent set and big enough to span the whole vector space. A basis of a finite dimensional vector space is not unique. However, every basis contains the same number of vectors. This number, called dimension, quantifies the size of the finite dimensional vector space. In the second case above, infinite dimensional vector spaces, we do not introduce here a concept of basis. More structure is needed in the vector space to be able to determine whether or not an infinite sum of vectors converges. We will not discuss these issues here.

4.3.1. **Basis of a vector space.** A particular type of finite sets in a vector space, small enough to be linearly independent and big enough to span the whole vector space, is called a basis of that vector space. Vector spaces having a finite set with these properties are essentially small, and they are called finite dimensional. When there is no finite set that spans the whole vector space, we call that space infinite dimensional. We now highlight these ideas in a more precise way.

Definition 4.3.1. *A finite set $\mathcal{S} \subset V$ is called a* ***finite basis*** *of a vector space V iff $\mathcal{S}$ is linearly independent and* $\text{Span}(\mathcal{S}) = V$.

The existence of a finite basis is the property that defines the size of the vector space.

Definition 4.3.2. *A vector space V is* ***finite dimensional*** *iff V has a finite basis or V is one of the following two extreme cases: $V = \emptyset$ or $V = \{\mathbf{0}\}$. Otherwise, the vector space V is called* ***infinite dimensional****.*

In these notes we will often call a finite basis just simply as a basis, without remarking that they contain a finite number of elements. We only study this type of basis, and we do not introduce the concept of an infinite basis. Why don't we define the notion of an infinite basis, since we have already defined the notion of an infinite linearly independent set? Because we do not have any way to define what is the span of an infinite set of vectors. In a vector space, without any further structure, there is no way to know whether an infinite sum converges or not. The notion of convergence needs further structure in a vector space, for example it needs a notion of distance between vectors. So, only in certain vector spaces with a notion of distance it is possible to introduce an infinite basis. We will discuss this subject in a later Chapter.

Example 4.3.1: We now present several examples.

(a) Let $V = \mathbb{R}^2$, then the set $\mathcal{S}_2 = \left\{ \mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ is a basis of $\mathbb{F}^2$. Notice that $\mathsf{e}_i = \mathsf{I}_{:i}$, that is, e_i is the i-th column of the identity matrix I_2. This basis $\mathcal{S}_2$ is called the standard basis of $\mathbb{R}^2$.

(b) A vector space can have infinitely many bases. For example, a second basis for $\mathbb{R}^2$ is the set $\mathcal{U} = \left\{ \mathsf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$. It is not difficult to verify that this set is a basis of $\mathbb{R}^2$, since u is linearly independent, and $\text{Span}(\mathcal{U}) = \mathbb{R}^2$.

(c) Let $V = \mathbb{F}^n$, then the set $\mathcal{S}_n = \left\{ \mathsf{e}_1 = \mathsf{I}_{:1}, \cdots, \mathsf{e}_n = \mathsf{I}_{:n} \right\}$ is a basis of $\mathbb{R}^n$, where $\mathsf{I}_{:i}$ is the i-th column of the identity matrix I_n. This set $\mathcal{S}_n$ is called the ***standard basis*** of $\mathbb{F}^n$.

(d) A basis for the vector space $\mathbb{F}^{2,2}$ of all 2×2 matrices is the set $\mathcal{S}_{2,2}$ given by

$$\mathcal{S}_{2,2} = \left\{ \mathsf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathsf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\};$$

This set is linearly independent and $\text{Span}(\mathcal{S}_{2,2}) = \mathbb{F}^{2,2}$, since any element $\mathsf{A} \in \mathbb{F}^{2,2}$ can be decomposed as follows,

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = A_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + A_{12} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + A_{21} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + A_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

(e) A basis for the vector space $\mathbb{F}^{m,n}$ of all $m \times n$ matrices is the following:

$$\mathcal{S}_{m,n} = \{ \mathsf{E}_{11}, \mathsf{E}_{12}, \cdots, \mathsf{E}_{mn} \},$$

where each $m \times n$ matrix E_{ij} is a matrix with all coefficients zero except the coefficient (i, j) which is equal to one (see previous example). The set $\mathcal{S}_{m,n}$ is linearly independent, and $\text{Span}(\mathcal{S}_{m,n}) = \mathbb{F}^{m,n}$, since

$$A = [A_{ij}] = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \mathsf{E}_{ij}.$$

(f) Let $V = \mathbb{P}_n$, the set of all polynomials with domain $\mathbb{R}$ and degree less or equal n. Any element $\boldsymbol{p} \in \mathbb{P}_n$ can be expressed as follows,

$$\boldsymbol{p}(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

that is equivalent to say that the set

$$\mathcal{S} = \{ \boldsymbol{p}_0 = 1, \boldsymbol{p}_1 = x, \boldsymbol{p}_2 = x^2, \cdots, \boldsymbol{p}_n = x^n \}$$

satisfies $\mathbb{P}_n = \text{Span}(\mathcal{S})$. The set $\mathcal{S}$ is also linearly independent, since

$$\boldsymbol{q}(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n = 0 \quad \Rightarrow \quad c_0 = \cdots = c_n = 0.$$

The proof of the latter statement is simple: Compute the n-th derivative of $\boldsymbol{q}$ above, and obtain the equation $n!\, c_n = 0$, so $c_n = 0$. Add this information into the $(n-1)$-th derivative of $\boldsymbol{q}$ and we conclude that $c_{n-1} = 0$. Continue in this way, and you will prove that all the coefficient c's vanish. Therefore, $\mathcal{S}$ is a basis of $\mathbb{P}_n$, ands it is also called the standard basis of $\mathbb{P}_n$.

◁

EXAMPLE 4.3.2: Show that the set $\mathcal{U} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$.

SOLUTION: We must show that $\mathcal{U}$ is a linearly independent set and $\text{Span}(\mathcal{U}) = \mathbb{R}^3$. Both properties follow from the fact that matrix U below, whose columns are the elements in $\mathcal{U}$, is invertible,

$$\mathsf{U} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{U}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & -1 \end{bmatrix}.$$

Let us show that $\mathcal{U}$ is a basis of $\mathbb{R}^3$: Since matrix U is invertible, this implies that its reduced echelon form $\mathsf{E}_{\mathsf{U}} = \mathsf{I}_3$, so its column vectors form a linearly independent set. The existence of U^{-1} implies that the system of equations $\mathsf{U}\mathsf{x} = \mathsf{y}$ has a solution $\mathsf{x} = \mathsf{U}^{-1}\mathsf{y}$ for every $\mathsf{y} \in \mathbb{R}^3$, that is, $\mathsf{y} \in \text{Col}(\mathsf{U}) = \text{Span}(\mathcal{U})$ for all $\mathsf{y} \in \mathbb{R}^3$. This means that $\text{Span}(\mathcal{U}) = \mathbb{R}^3$. Hence, the set $\mathcal{U}$ is a basis of $\mathbb{R}^3$. ◁

The following definitions will be useful to establish important properties of a basis.

Definition 4.3.3. *Let V be a vector space and $S_n \subset V$ be a subset with n elements. The set S_n is a* ***maximal linearly independent*** *set iff S_n is linearly independent and every other set $\tilde{S}_m$ with $m > n$ elements is linearly dependent. The set S_n is a* ***minimal spanning*** *set iff* $\text{Span}(S_n) = V$ *and every other set $\tilde{S}_m$ with $m < n$ elements satisfies* $\text{Span}(\tilde{S}_m) \subsetneq V$.

A maximal linearly independent set S is the biggest set in a vector space that is linearly independent. A set cannot be linearly independent if it is too big, since the bigger the set the more probable that one element in the set is a linear combination of the other elements in the set. A minimal spanning set is the smallest set in a vector space that spans the whole space. A spanning set, that is, a set whose span is the whole space, cannot be too small, since the smaller the set the more probable that an element in the vector space is outside the span of the set. The following result provides a useful characterization of a basis: *A basis is a set in the vector space that is both maximal linearly independent and minimal spanning.* In this sense, a basis is a set with the right size, small enough to be linearly independent and big enough to span the whole vector space.

Theorem 4.3.4. *Let V be a vector space. The following statements are equivalent:*

(a) $\mathcal{U}$ is a basis of V;
(b) $\mathcal{U}$ is a minimal spanning set in V.
(c) $\mathcal{U}$ is a maximal linearly independent set in V;

EXAMPLE 4.3.3: We showed in Example 4.3.2 above that the set $\mathcal{U} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$. Since this basis has three elements, Theorem 4.3.4 says that any other spanning set in $\mathbb{R}^3$ cannot have less than three vectors, and any other linearly independent set in $\mathbb{R}^3$ cannot have more that three vectors. For example, any subset of $\mathcal{U}$ containing two elements cannot span $\mathbb{R}^3$; the linear combination of two vectors in $\mathcal{U}$ span a plane in $\mathbb{R}^3$. Another example, any set of four vectors in $\mathbb{R}^3$ must be linearly dependent. ◁

Proof of Theorem 4.3.4: We first show part (a)-(b).

(⇒) Assume that $\mathcal{U}$ is a basis of V. If the set $\mathcal{U}$ is not a minimal spanning set of V, that means there exists $\tilde{\mathcal{U}} = \{\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_{n-1}\}$ such that $\text{Span}(\tilde{\mathcal{U}}) = V$. So, every vector in $\mathcal{U}$ can be expressed as a linear combination of vectors in $\tilde{\mathcal{U}}$. Hence, there exists a set of coefficients C_{ij} such that

$$\boldsymbol{u}_j = \sum_{i=1}^{n-1} C_{ij} \tilde{\boldsymbol{u}}_i, \qquad j = 1, \cdots, n.$$

The reason to order the coefficients C_{ij} is this form is that they form a matrix $\mathsf{C} = [C_{ij}]$ which is $(n-1) \times n$. This matrix C defines a function $\mathsf{C} : \mathbb{R}^n \to \mathbb{R}^{n-1}$, and since $\text{rank}(\mathsf{C}) \leqslant (n-1) < n$, this matrix satisfies that $N(\mathsf{C})$ is nontrivial as a subset of $\mathbb{R}^n$. So there exists a nonzero column vector in $\mathbb{R}^n$ with components $\mathsf{z} = [z_j] \in \mathbb{R}^n$, not all components zero, such that $\mathsf{z} \in N(\mathsf{C})$, that is,

$$\sum_{j=1}^{n} C_{ij} z_j = 0, \qquad i = 1, \cdots, (n-1).$$

What we have found is that the linear combination

$$z_1 \boldsymbol{u}_1 + \cdots + z_n \boldsymbol{u}_n = \sum_{j=1}^{n} z_j \boldsymbol{u}_j = \sum_{j=1}^{n} z_j \big(\sum_{i=1}^{n-1} C_{ij} \tilde{\boldsymbol{u}}_i \big) = \sum_{i=1}^{n-1} \big(\sum_{j=1}^{n} C_{ij} z_j \big) \tilde{\boldsymbol{u}}_i = \boldsymbol{0},$$

with at least one of the coefficients z_j non-zero. This means that the set $\mathcal{U}$ is not linearly independent. But this contradicts that $\mathcal{U}$ is a basis. Therefore, the set $\mathcal{U}$ is a minimal spanning set of V.

($\Leftarrow$) Assume that $\mathcal{U}$ is a minimal spanning set of V. If $\mathcal{U}$ is not a basis, that means $\mathcal{U}$ is not a linearly independent set. At least one element in $\mathcal{U}$ must be a linear combination of the others. Let us arrange the order of the basis vectors such that the vector $\boldsymbol{u}_n$ is a linear combination of the other vectors in $\mathcal{U}$. Then, the set $\tilde{\mathcal{U}} = \{\boldsymbol{u}, \cdots, \boldsymbol{u}_{n-1}\}$ must still span V, that is $\text{Span}(\tilde{\mathcal{U}}) = V$. But this contradicts the assumption that $\mathcal{U}$ is a minimal spanning set of V.

We now show part (a)-(c).

($\Leftarrow$) Assume that $\mathcal{U}$ is a maximal linearly independent set in V. If $\mathcal{U}$ is not a basis, that means $\text{Span}(\mathcal{U}) \subsetneqq V$, so there exists $\boldsymbol{u}_{n+1} \in V$ such that $\boldsymbol{u}_{n+1} \notin \text{Span}(\mathcal{U})$. Hence, the set $\tilde{\mathcal{U}} = \{\boldsymbol{u}, \cdots, \boldsymbol{u}_{n+1}\}$ is a linearly independent set. However, this contradicts the assumption that $\mathcal{U}$ is a maximal linearly independent set. We conclude that $\mathcal{U}$ is a basis of V.

($\Rightarrow$) Assume that $\mathcal{U}$ is a basis of V. If the set $\mathcal{U}$ is not a maximal linearly independent set in V, then there exists a maximal linearly independent set $\tilde{\mathcal{U}} = \{\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_k\}$, with $k > n$. By the argument given just above, $\tilde{\mathcal{U}}$ is a basis of V. By part (b) the set $\tilde{\mathcal{U}}$ must be a minimal spanning set of V. However, this is not true, since $\mathcal{U}$ is smaller and spans V. Therefore, $\mathcal{U}$ must be a maximal linearly independent set in V.

This establishes the Theorem. □

4.3.2. **Dimension of a vector space.** The characterization of a basis given in Theorem 4.3.4 above implies that the number of elements in a basis is always the same as in any other basis.

Theorem 4.3.5. *The number of elements in any basis of a finite dimensional vector space is the same as in any other basis.*

Proof of Theorem (4.3.5: Let $\mathcal{V}_n$ and $\mathcal{V}_m$ be two bases of a vector space V with n and m elements, respectively. If $m > n$, the property that $\mathcal{V}_m$ is a minimal spanning set implies that $\text{Span}(\mathcal{V}_n) \subsetneqq \text{Span}(\mathcal{V}_m) = V$. The former inclusion contradicts that $\mathcal{V}_n$ is a basis. Therefore, $n = m$. (A similar proof can be constructed with the maximal linearly independence property of a basis.) This establishes the Theorem. □

The number of elements in a basis of a finite dimensional vector space is a characteristic of the vector space, so we give that characteristic a name.

Definition 4.3.6. *The **dimension** of a finite dimensional vector space V with a finite basis, denoted as $\dim V$, is the number of elements in any basis of V. The extreme cases of $V = \emptyset$ and $V = \{\mathbf{0}\}$ are defined as zero dimensional.*

From the definition above we see that $\dim\{\boldsymbol{0}\} = 0$ and $\dim \emptyset = 0$.

EXAMPLE 4.3.4: We now present several examples.

(a) The set $\mathcal{S}_n = \{\mathsf{e}_1 = \mathsf{I}_{:1}, \cdots, \mathsf{e}_n = \mathsf{I}_{:n}\}$ is a basis for $\mathbb{F}^n$, so $\dim \mathbb{F}^n = n$.

(b) A basis for the vector space $\mathbb{F}^{2,2}$ of all 2×2 matrices is the set $\mathcal{S}_{2,2}$ is given by

$$\mathcal{S}_{2,2} = \left\{\mathsf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathsf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right\},$$

so we conclude that $\dim \mathbb{F}^{2,2} = 4$.

(c) A basis for the vector space $\mathbb{F}^{m,n}$ of all $m \times n$ matrices is the following:

$$\mathcal{S}_{m,n} = \{\mathsf{E}_{11}, \mathsf{E}_{12}, \cdots, \mathsf{E}_{mn}\},$$

where we recall that each $m \times n$ matrix E_{ij} is a matrix with all coefficients zero except the coefficient (i, j) which is equal to one. Since the basis $\mathcal{S}_{m,n}$ contains mn elements, we conclude that $\dim \mathbb{F}^{m,n} = mn$.

(d) A basis for the vector space $\mathbb{P}_n$ of all polynomial with degree less or equal n is the set $\mathcal{S}$ given by $\mathcal{S} = \{\boldsymbol{p}_0 = 1, \boldsymbol{p}_1 = x, \boldsymbol{p}_2 = x^2, \cdots, \boldsymbol{p}_n = x^n\}$. This set has $n + 1$ elements, so $\dim \mathbb{P}_n = n + 1$.

◁

REMARK: Any subspace $W \subset V$ of a vector space V is itself a vector space, so the definition of basis also holds for W. Since $W \subset V$, we conclude that $\dim W \leqslant \dim V$.

EXAMPLE 4.3.5: Consider the case $V = \mathbb{R}^3$. It is simple to see in Fig. 37 that $\dim U = 1$ and $\dim W = 2$, where the subspaces U and W are spanned by one vector and by two non-collinear vectors, respectively.

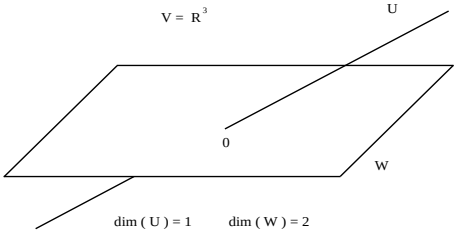

FIGURE 37. Sketch of two subspaces U and W, in the vector space $\mathbb{R}^3$, of dimension one and two, respectively.

◁

EXAMPLE 4.3.6: Find a basis for $N(\mathsf{A})$ and $R(\mathsf{A})$, where matrix $\mathsf{A} \in \mathbb{R}^{3,4}$ is given by

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}. \tag{4.2}$$

SOLUTION: Since $\mathsf{A} \in \mathbb{R}^{3,4}$, then $\mathsf{A} : \mathbb{R}^4 \to \mathbb{R}^3$, which implies that $N(\mathsf{A}) \subset \mathbb{R}^4$ while $R(\mathsf{A}) \subset \mathbb{R}^3$. A basis for $N(\mathsf{A})$ is found as follows: Find all solution of $\mathsf{Ax} = \mathsf{0}$ and express these solutions as the span of a linearly independent set of vectors. We first find E_A,

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & \frac{5}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathsf{E}_\mathsf{A} \quad \Rightarrow \quad \begin{cases} x_1 = -6x_3 - 5x_4, \\ x_2 = -\dfrac{5}{2}\, x_3 - \dfrac{3}{2}\, x_4, \\ x_3,\ x_4 \quad \text{free variables.} \end{cases}$$

Therefore, every element in $N(\mathsf{A})$ can be expressed as follows,

$$\mathsf{x} = \begin{bmatrix} -6x_3 - 5x_4 \\ -\frac{5}{2}\, x_3 - \frac{3}{2}\, x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} x_4, \quad \Rightarrow \quad N(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\} \right).$$

Since the vectors in the span above form a linearly independent set, we conclude that a basis for $N(\mathsf{A})$ is the set $\mathcal{N}$ given by

$$\mathcal{N} = \left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We now find a basis for $R(\mathsf{A})$. We know that $R(\mathsf{A}) = \text{Col}(\mathsf{A})$, that is, the span of the column vectors of matrix A. We only need to find a linearly independent subset of column vectors of A. This information is given in $\mathsf{E_A}$, since the pivot columns in $\mathsf{E_A}$ indicate the columns in A which form a linearly independent set. In our case, the pivot columns in $\mathsf{E_A}$ are the first and second columns, so we conclude that a basis for $R(\mathsf{A})$ is the set $\mathcal{R}$ given by

$$\mathcal{R} = \left\{ \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} \right\}.$$

◁

4.3.3. **Extension of a set to a basis.** We know that a basis of a vector space is not unique, and the following result says that actually any linearly independent set can be extended into a basis of a vector space.

Theorem 4.3.7. *If $S_k = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k\}$ is a linearly independent set in a vector spave V with basis $\mathcal{V} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$, where $k < n$, then, there always exists a basis of V given by an extension of the set S_k of the form $\mathcal{S} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k, \boldsymbol{v}_{i_1}, \cdots, \boldsymbol{v}_{i_{n-k}}\}$.*

The statement above says that a linearly independent set S_k can be extended into a basis $\mathcal{S}$ of a a vector space V simply incorporating appropriate vectors from any basis of V. If a basis $\mathcal{V}$ of V has n vectors, and the set S_k has $k < n$ vectors, then one can always select $n - k$ vectors from the basis $\mathcal{V}$ to enlarge the set S_k into a basis of V.

Proof of Theorem 4.3.7: Introduce the set S_{k+n}

$$s_{k+n} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}.$$

We know that $\text{Span}(S_{k+n}) = V$ since $\mathcal{V} \subset S_{k+n}$. We also know that S_{k+n} is linearly dependent, since the maximal linearly independent set contains n elements and S_{k+n} contains $n + k > n$ elements. The idea is to eliminate the $\boldsymbol{v}_i$ such that $S_k \cup \{\boldsymbol{v}_i\}$ is linearly dependent. Since the maximal linearly independent set contains n elements and the S_k is linearly independent, there are k elements in $\mathcal{V}$ that will be eliminated. The resulting set is $\mathcal{S}$, which is a basis of V containing S_k. This establishes the Theorem. □

EXAMPLE 4.3.7: Given the 3×4 matrix A defined in Eq. (4.2) in Example 4.3.6 above, extend the basis of $N(\mathsf{A}) \subset \mathbb{R}^4$ into a basis of $\mathbb{R}^4$.

SOLUTION: We know from Example 4.3.6 that a basis for the $N(\mathsf{A})$ is the set $\mathcal{N}$ given by

$$\mathcal{N} = \left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Following the idea in the proof of Theorem 4.3.7, we look for a linear independent set of vectors among the columns of the matrix

$$\mathsf{M} = \begin{bmatrix} -6 & -5 & 1 & 0 & 0 & 0 \\ -\frac{5}{2} & -\frac{3}{2} & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

That is, matrix M include the basis vectors of $N(\mathsf{A})$ and the four vectors e_i of the standard basis of $\mathbb{R}^4$. It is important to place the basis vectors of $N(\mathsf{A})$ in the first columns of M. In this way, the Gauss method will select these first vectors as part of the linearly independent set of vectors. Find now the reduced echelon form matrix E_M,

$$\mathsf{E}_\mathsf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 6 & 5 \\ 0 & 0 & 0 & 1 & 5 & 3 \end{bmatrix}.$$

Therefore, the first four vectors in M are form a linearly independent set, so a basis of $\mathbb{R}^4$ that includes $\mathcal{N}$ is given by

$$\mathcal{V} = \left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

◁

4.3.4. **The dimension of subspace addition.** Recall that the sum and the intersection of two subspaces is again a subspace in a given vector space. The following result relates the dimension of a sum of subspaces with the dimension of the individual subspaces and the dimension of their intersection.

Theorem 4.3.8. *If W_1, $W_2 \subset V$ are subspaces of a vector space V, then holds*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Proof of Theorem 4.3.8: We find the dimension of $W_1 + W_2$ finding a basis of this sum. The key idea is to start with a basis of $W_1 \cap W_2$. Let $\mathcal{B}_0 = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l\}$ be a basis for $W_1 \cap W_2$. Enlarge that basis into basis $\mathcal{B}_1$ for W_1 and $\mathcal{B}_2$ for W_2 as follows,

$$\mathcal{B}_1 = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}, \quad \mathcal{B}_2 = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}.$$

We use the notation $l = \dim(W_1 \cap W_2)$, $l + n = \dim W_1$ and $l + m = \dim W_2$. We now propose as basis for $W_1 + W_2$ the set

$$\mathcal{B} = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}.$$

By construction this set satisfies that $\text{Span}(\mathcal{B}) = W_1 + W_2$. We only need to show that $\mathcal{B}$ is linearly independent. Assume that the set $\mathcal{B}$ is linearly dependent. This means that there is non-zero constants a_i, b_j and c_k solutions of the equation

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i + \sum_{j=1}^{m} b_j \boldsymbol{y}_j + \sum_{k=1}^{l} c_k \boldsymbol{z}_k = \boldsymbol{0}. \tag{4.3}$$

This implies that the vector $\sum_{i=1}^{n} a_i \boldsymbol{x}_i$, which by definition belongs to W_1, also belongs to W_2, since

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i = -\Big(\sum_{j=1}^{m} b_j \boldsymbol{y}_j + \sum_{k=1}^{l} c_k \boldsymbol{z}_k\Big) \in W_2.$$

Therefore, $\sum_{i=1}^{n} a_i \boldsymbol{x}_i$ belongs to $W_1 \cap W_2$, and so is a linear combination of the elements of $\mathcal{B}_0$, that is, there exists scalars d_k such that

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i = \sum_{i=1}^{l} d_k \boldsymbol{z}_k.$$

Since $\mathcal{B}_1$ is a basis of W_1, this implies that all the coefficients a_i and d_k vanish. Introduce this information into Eq. (4.3) and we conclude that

$$\sum_{j=1}^{m} b_j \boldsymbol{y}_j + \sum_{k=1}^{l} c_k \boldsymbol{z}_k = \boldsymbol{0}.$$

Analogously, the set $\mathcal{B}_2$ is a basis, so all the coefficients b_j and c_k must vanish. This implies that the set $\mathcal{B}$ is linearly independent, hence a basis of $W_1 + W_2$. Therefore, the dimension of the sum is given by

$$\dim(W_1 + W_2) = n + m + k = (n + k) + (m + k) - k = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

This establishes the Theorem. □

The following corollary is immediate.

Corollary 4.3.9. *If a vector space can be decomposed as $V = W_1 \oplus W_2$, then*

$$\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2.$$

The proof is straightforward from Theorem 4.3.8, since the condition of subspaces direct sum, $W_1 \cap W_2 = \{\boldsymbol{0}\}$, says that $\dim\big(W_1 \cap W_2\big) = 0$.

4.3.5. **Exercises.**

4.3.1.- Find a basis for each of the spaces $N(\mathsf{A})$, $R(\mathsf{A})$, $N(\mathsf{A}^T)$, $R(\mathsf{A}^T)$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{bmatrix}.$$

4.3.2.- Find the dimension of the space spanned by

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \\ -4 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 0 \\ 6 \end{bmatrix} \right\}.$$

4.3.3.- Find the dimension of the following spaces:

(a) The space $\mathbb{P}_n$ of polynomials of degree less or equal n.

(b) The space $\mathbb{F}^{m,n}$ of $m \times n$ matrices.

(c) The space of real symmetric $n \times n$ matrices.

(d) The space of real skew-symmetric $n \times n$ matrices.

4.3.4.- Find an example to show that the following statement is false: Given a basis $\{\mathsf{v}_1, \mathsf{v}_2\}$ of $\mathbb{R}^2$, then every subspace $W \subset \mathbb{R}^2$ has a basis containing at least one of the basis vectors v_1, v_2.

4.3.5.- Given the matrix A and vector v,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 2 & 0 & 5 \\ 2 & 4 & 3 & 1 & 8 \\ 3 & 6 & 1 & 5 & 5 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} -8 \\ 1 \\ 3 \\ 3 \\ 0 \end{bmatrix},$$

verify that $\mathsf{v} \in N(\mathsf{A})$, and then find a basis of $N(\mathsf{A})$ containing v.

4.3.6.- Determine whether or not the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

is a basis for the subspace

$$\text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 8 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \right\} \right) \subset \mathbb{R}^3.$$

4.4. Vector components

4.4.1. **Ordered bases.** In the previous Section we introduced a finite basis in a vector space. Although a vector space can have different bases, every basis has the same number of elements, which provides a measure of the vector space size, called the dimension of the vector space. In this Section we study another property of a basis. Every vector in a finite dimensional vector space can be expressed in a unique way as a linear combination of the basis vectors. This property can be clearly stated in an ordered basis, which is a basis with the basis vectors given in a specific order.

Definition 4.4.1. *An* ***ordered basis*** *of an n-dimensional vector space V is a sequence $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ of vectors such that the set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ is a basis of V.*

Recall that a sequence is an ordered set, that is, a set with elements given in a particular order.

Example 4.4.1: The following four ordered basis of $\mathbb{R}^3$ are all different,

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right), \quad \left(\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right), \quad \left(\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right), \quad \left(\begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right),$$

however, they determine the same basis $\mathcal{S}_3 = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$. ◁

4.4.2. **Vector components in a basis.** The following result states that given a vector space with an ordered basis, there exists a correspondence between vectors and certain sequences of scalars.

Theorem 4.4.2. *Let V be an n-dimensional vector space over the scalar field $\mathbb{F}$ with an ordered basis $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$. Then, every vector $\boldsymbol{v} \in V$ determines a unique scalars' sequence $(v_1, \cdots, v_n) \subset \mathbb{F}$ such that*

$$\boldsymbol{v} = v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n. \tag{4.4}$$

And every scalars' sequence $(v_1, \cdots, v_n) \subset \mathbb{F}$ determines a unique vector $\boldsymbol{v} \in V$ by Eq. (4.4).

Proof of Theorem 4.4.2: Denote by $\mathcal{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ the ordered basis of V. Since $\mathcal{U}$ is a basis, $\text{Span}(\mathcal{U}) = V$ and $\mathcal{U}$ is linearly independent. The first property implies that for every $\boldsymbol{v} \in V$ there exist scalars $v_1, \cdots, v_n$ such that $\boldsymbol{v}$ is a linear combination of the basis vectors, that is,

$$\boldsymbol{v} = v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n.$$

The second property of a basis implies that the linear combination above is unique. Indeed, if there exists another linear combination

$$\boldsymbol{v} = \nu_1 \boldsymbol{u}_1 + \cdots + \nu_n \boldsymbol{u}_n,$$

then $\boldsymbol{0} = \boldsymbol{v} - \boldsymbol{v} = (v_1 - \nu_1)\boldsymbol{u}_1 + \cdots + (v_n - \nu_n)\boldsymbol{u}_n$. Since $\mathcal{U}$ is linearly independent, this implies that each coefficient above vanishes, so $v_1 = \nu_1$, $\cdots$, $v_n = \nu_n$.

The converse statement is simple to show, since the scalars are given in a specific order. Every scalars' sequence $(v_1, \cdots, v_n)$ determines a unique linear combination with an ordered basis $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ given by $v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n$. This unique linear combination determines a unique vector in the vector space. This establishes the Theorem. □

Theorem 4.4.2 says that there exists a correspondence between vectors in a vector space with an ordered basis and scalars' sequences. This correspondence is called a coordinate map and the scalars are called vector components in the basis. Here is a precise definition.

Definition 4.4.3. *Let V be an n-dimensional vector space over $\mathbb{F}$ with an ordered basis $\mathcal{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$. The* ***coordinate map*** *is the function $[\ ]_u : V \to \mathbb{F}^n$, with $[\boldsymbol{v}]_u = \mathsf{v}_u$, and*

$$\mathsf{v}_u = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{v} = v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n .$$

The scalars $v_1, \cdots, v_n$ are called the ***vector components*** *of $\boldsymbol{v}$ in the ordered basis $\mathcal{U}$.*

Therefore, we use the notation $[\boldsymbol{v}]_u = \mathsf{v}_u \in \mathbb{F}^n$ for the components of a vector $\boldsymbol{v} \in V$ in an ordered basis $\mathcal{U}$. We remark that the coordinate map is defined only after an ordered basis is fixed in V. Different ordered bases on V determine different coordinate maps between V and $\mathbb{F}^n$. When the situation under study involves only one ordered basis, we suppress the basis subindex. The coordinate map will be denoted by $[\ ] : V \to \mathbb{F}^n$ and the vector components by $\mathsf{v} = [\boldsymbol{v}]$. In the particular case that $V = \mathbb{F}^n$ and the basis is the standard basis $\mathcal{S}_n$, then the coordinate map $[\ ]_s$ is the identity map, so $\boldsymbol{v} = [\boldsymbol{v}]_s = \mathsf{v}$. In this case we follow the convention established in the first Chapters, that is, we denote vectors in $\mathbb{F}^n$ by v instead of $\boldsymbol{v}$. When the situation under study involves more than one ordered basis we keep the sub-indices in the coordinate map, like $[\ ]_u$, and in the vector components, like v_u, to keep track of the basis attached to these expressions.

EXAMPLE 4.4.2: Let V be the set of points on the plane with a preferred origin. Let $\mathcal{S} = \big(\boldsymbol{e}_1,\ \boldsymbol{e}_2\big)$ be an ordered basis, pictured in Fig. 38.

(a) Find the components $\mathsf{v}_s = [\boldsymbol{v}]_s \in \mathbb{R}^2$ of the vector $\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2$ in the ordered basis $\mathcal{S}$.
(b) Find the components $\mathsf{v}_u = [\boldsymbol{v}]_u \in \mathbb{R}^2$ of the same vector $\boldsymbol{v}$ given in part (a) but now in the ordered basis $\mathcal{U} = \big(\boldsymbol{u}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2,\ \boldsymbol{u}_2 = -\boldsymbol{e}_1 + \boldsymbol{e}_2\big)$.

SOLUTION: Part (a) is straightforward to compute, since the definition of component of a vector says that the numbers multiplying the basis vectors in the equation $\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2$ are the components of the vector, that is,

$$\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2 \quad \Leftrightarrow \quad \mathsf{v}_s = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Part (b) is more involved. We are looking for numbers $\tilde{v}_1$ and $\tilde{v}_2$ such that

$$\boldsymbol{v} = \tilde{v}_1 \boldsymbol{u}_1 + \tilde{v}_2 \boldsymbol{u}_2 \quad \Leftrightarrow \quad \mathsf{v}_u = \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix}. \tag{4.5}$$

From the definition of the basis $\mathcal{U}$ we know the components of the basis vectors in $\mathcal{U}$ in terms of the standard basis, that is,

$$\boldsymbol{u}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2 \quad \Leftrightarrow \quad \mathsf{u}_{1s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\boldsymbol{u}_2 = -\boldsymbol{e}_1 + \boldsymbol{e}_2 \quad \Leftrightarrow \quad \mathsf{u}_{2s} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

In other words, we can write the ordered basis $\mathcal{U}$ as the column vectors of the matrix $\mathsf{U}_s = \big[\mathcal{U}\big]_s = \big[[\boldsymbol{u}_1]_s, [\boldsymbol{u}_2]_s\big]$ given by

$$\mathsf{U}_s = \big[\mathsf{u}_{1s}, \mathsf{u}_{2s}\big] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Expressing Eq. (4.5) in the standard basis means

$$\boldsymbol{e}_1 + 3\boldsymbol{e}_2 = \boldsymbol{v} = \tilde{v}_1(\boldsymbol{e}_1 + \boldsymbol{e}_2) + \tilde{v}_2(-\boldsymbol{e}_1 + \boldsymbol{e}_2) \quad \Leftrightarrow \quad \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \tilde{v}_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \tilde{v}_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The last equation on the right is a matrix equation for the unknowns $\mathsf{v}_u = \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix}$,

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \Leftrightarrow \quad \mathsf{U}_s \mathsf{v}_u = \mathsf{v}_s.$$

We find the solution using the Gauss method,

$$\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 1 & 1 & 3 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \end{array}\right] \quad \Rightarrow \quad \mathsf{v}_u = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{v} = 2\boldsymbol{u}_1 + \boldsymbol{u}_2.$$

A sketch of what has been computed is in Fig. 38. In this Figure is clear that the vector $\boldsymbol{v}$ is fixed, and we have only expressed this fixed vector in as a linear combination of two different bases. It is clear in this Fig. 38 that one has to stretch the vector $\boldsymbol{u}_1$ by two and add the result to the vector $\boldsymbol{u}_2$ to obtain $\boldsymbol{v}$. ◁

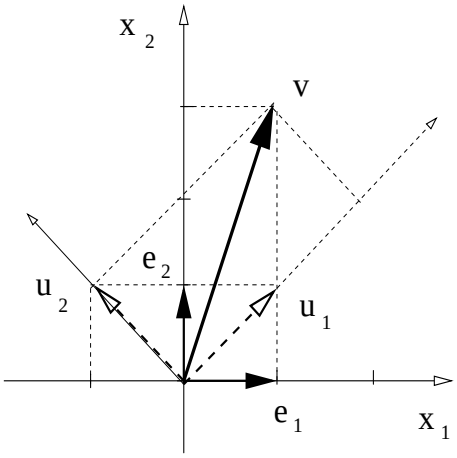

FIGURE 38. The vector $\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2$ expressed in terms of the basis $\mathcal{U} = \{\boldsymbol{u}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2, \boldsymbol{u}_2 = -\boldsymbol{e}_1 + \boldsymbol{e}_2\}$ is given by $\boldsymbol{v} = 2\boldsymbol{u}_1 + \boldsymbol{u}_2$.

EXAMPLE 4.4.3: Consider the vector space $\mathbb{P}_2$ of all polynomials of degree less or equal two, and let us consider the case of $\mathbb{F} = \mathbb{R}$. An ordered basis is $\mathcal{S} = (\boldsymbol{p}_0 = 1, \boldsymbol{p}_1 = x, \boldsymbol{p}_2 = x^2)$. The coordinate map is $[\,]_s : \mathbb{P}_2 \to \mathbb{R}^3$ defined as follows, $[\boldsymbol{p}]_s = \mathsf{p}_s$, where

$$\mathsf{p}_s = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{p}(x) = a + bx + cx^2.$$

The column vector p_s represents the components of the vector $\boldsymbol{p}$ in the ordered basis $\mathcal{S}$. The equation above defines a correspondence between every element in $\mathbb{P}_2$ and every element in $\mathbb{R}^3$. The coordinate map depend on the choice of the ordered basis. For example, choosing the ordered basis $\tilde{\mathcal{S}} = (\boldsymbol{p}_0 = x^2, \boldsymbol{p}_1 = x, \boldsymbol{p}_2 = 1)$, the corresponding coordinate map is $[\,]_{\tilde{s}} : \mathbb{P}_2 \to \mathbb{R}^3$ defined by $[\boldsymbol{p}]_{\tilde{s}} = \mathsf{p}_{\tilde{s}}$, where

$$\mathsf{p}_{\tilde{s}} = \begin{bmatrix} c \\ b \\ a \end{bmatrix} \quad \Leftrightarrow \quad p(x) = a + bx + cx^2.$$

The coordinate maps above generalize to the spaces $\mathbb{P}_n$ and $\mathbb{R}^{n+1}$ for all $n \in \mathbb{N}$. Given the ordered basis $\mathcal{S} = \big(\boldsymbol{p}_0 = 1, \boldsymbol{p}_1 = x, \cdots, \boldsymbol{p}_n = x^n\big)$, the corresponding coordinate map $[\,]_s : \mathbb{P}_n \to \mathbb{R}^{n+1}$ defined by $[\boldsymbol{p}]_s = \mathsf{p}_s$, where

$$\mathsf{p}_s = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} \quad \Leftrightarrow \quad p(x) = a_0 + \cdots + a_n x^n.$$

◁

EXAMPLE 4.4.4: Consider $V = \mathbb{P}_2$ with ordered basis $\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2\big)$.

(a) Find $\mathsf{r}_s = [\boldsymbol{r}]_s$, the components of $\boldsymbol{r}(x) = 3 + 2x + 4x^2$ in the ordered basis $\mathcal{S}$.
(b) Find $\mathsf{r}_q = [\boldsymbol{r}]_q$, the components of the same polynomial $\boldsymbol{r}$ given in part (a) but now in the ordered basis $\mathcal{Q} = \big(\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = 1 + x,\ \boldsymbol{q}_2 = 1 + x + x^2,\big)$.

SOLUTION:

Part (a):This is straightforward to compute, since $\boldsymbol{r}(x) = 3 + 2x + 4x^2$ implies that

$$\boldsymbol{r}(x) = 3\,\boldsymbol{p}_0(x) + 2\,\boldsymbol{p}_1(x) + 4\boldsymbol{p}_2(x) \quad \Leftrightarrow \quad \mathsf{r}_s = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}.$$

Part (b): This is more involved, as in Example 4.4.2. We look for numbers $\tilde{r}_1$, $\tilde{r}_2$, $\tilde{r}_3$ such that

$$\boldsymbol{r}(x) = \tilde{r}_0\,\boldsymbol{q}_0(x) + \tilde{r}_1\,\boldsymbol{q}_1(x) + \tilde{r}_2\,\boldsymbol{q}_2(x) \quad \Leftrightarrow \quad \mathsf{r}_q = \begin{bmatrix} \tilde{r}_0 \\ \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix}. \tag{4.6}$$

From the definition of the basis $\mathcal{Q}$ we know the components of the basis vectors in $\mathcal{Q}$ in terms of the $\mathcal{S}$ basis, that is,

$$\boldsymbol{q}_0(x) = \boldsymbol{p}_0(x) \qquad \Leftrightarrow \qquad \mathsf{q}_{0s} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\boldsymbol{q}_1(x) = \boldsymbol{p}_0(x) + \boldsymbol{p}_1(x) \qquad \Leftrightarrow \qquad \mathsf{q}_{1s} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\boldsymbol{q}_2(x) = \boldsymbol{p}_0(x) + \boldsymbol{p}_1(x) + \boldsymbol{p}_2(x) \qquad \Leftrightarrow \qquad \mathsf{q}_{2s} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Now we can write the ordered basis $\mathcal{Q}$ in terms of the column vectors of the matrix $\mathsf{Q}_s = \big[\mathcal{Q}\big]_s = \big[\mathsf{q}_{0s},\, \mathsf{q}_{1s},\, \mathsf{q}_{2s}\big]$, as follows,

$$\mathsf{Q}_s = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Expressing Eq. (4.6) in the standard basis means

$$\begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = \tilde{r}_0 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \tilde{r}_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \tilde{r}_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The last equation on the right is a matrix equation for the unknowns $\tilde{r}_0$, $\tilde{r}_1$, and $\tilde{r}_2$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{r}_0 \\ \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} \quad \Leftrightarrow \quad \mathsf{Q}_s \mathsf{r}_q = \mathsf{r}_s.$$

We find the solution using the Gauss method,

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 4 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 4 \end{array}\right],$$

hence the solution is

$$\mathsf{r}_q = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{r}(x) = \boldsymbol{q}_0(x) - 2\,\boldsymbol{q}_1(x) + 4\,\boldsymbol{q}_2(x).$$

We can verify that this is the solution, since

$$\begin{aligned} \boldsymbol{r}(x) &= \boldsymbol{q}_0(x) - 2\,\boldsymbol{q}_1(x) + 4\,\boldsymbol{q}_2(x) \\ &= 1 - 2\,(1 + x) + 4\,(1 + x + x^2) \\ &= (1 - 2 + 4) + (-2 + 4)x + 4x^2 \\ &= 3 + 2x + 4x^2 \\ &= 3\,\boldsymbol{p}_0(x) + 2\,\boldsymbol{p}_1(x) + 4\boldsymbol{p}_2(x). \end{aligned}$$

◁

EXAMPLE 4.4.5: Given any ordered basis $\mathcal{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ of a 3-dimensional vector space V, find $\mathsf{U}_u = \big[\mathcal{U}\big]_u \subset \mathbb{F}^{3,3}$, that is, find $\mathsf{u}_i = [\boldsymbol{u}_i]_u$ for $i = 1, 2, 3$, the components of the basis vectors $\boldsymbol{u}_i$ in its own basis $\mathcal{U}$.

SOLUTION: The answer is simple: The definition of vector components in a basis says that

$$\boldsymbol{u}_1 = \boldsymbol{u}_1 + 0\,\boldsymbol{u}_2 + 0\,\boldsymbol{u}_3 \quad \Leftrightarrow \quad \mathsf{u}_{1u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathsf{e}_1,$$

$$\boldsymbol{u}_2 = 0\,\boldsymbol{u}_1 + \boldsymbol{u}_2 + 0\,\boldsymbol{u}_3 \quad \Leftrightarrow \quad \mathsf{u}_{2u} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \mathsf{e}_2,$$

$$\boldsymbol{u}_3 = 0\,\boldsymbol{u}_1 + 0\,\boldsymbol{u}_2 + \boldsymbol{u}_3 \quad \Leftrightarrow \quad \mathsf{u}_{3u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathsf{e}_3.$$

In other words, using the coordinate map $\phi_u : V \to \mathbb{F}^3$, we can always write any basis $\mathcal{U}$ as components in its own basis as follows, $\mathsf{U}_u = \big[\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3\big] = \mathsf{I}_3$. This example says that there is nothing special about the standard basis $\mathcal{S} = \{\mathsf{e}_1, \cdots, \mathsf{e}_n\}$ of $\mathbb{F}^n$, where $\mathsf{e}_i = \mathsf{I}_{:i}$ is the i-th column of the identity matrix I_n. Given any n-dimensional vector space V over $\mathbb{F}$ with any ordered basis $\mathcal{V}$, the components of the basis vectors expressed on its own basis is always the standard basis of $\mathbb{F}^n$, that is, the result is always $\big[\mathcal{V}\big]_v = \mathsf{I}_n$. ◁

4.4.3. **Exercises.**

4.4.1.- Let $\mathcal{S} = \big(\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3\big)$ be the standard basis of $\mathbb{R}^3$. Find the components of the vector $\mathsf{v} = \mathsf{e}_1 + \mathsf{e}_2 + 2\mathsf{e}_3$ in the ordered basis $\mathcal{U}$

$$\Big(\mathsf{u}_{1s} = \begin{bmatrix}1\\0\\1\end{bmatrix}, \mathsf{u}_{2s} = \begin{bmatrix}0\\1\\1\end{bmatrix}, \mathsf{u}_{3s} = \begin{bmatrix}1\\1\\0\end{bmatrix}\Big).$$

4.4.2.- Let $\mathcal{S} = \big(\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3\big)$ be the standard basis of $\mathbb{R}^3$. Find the components of the vector

$$\mathsf{v}_s = \begin{bmatrix}8\\7\\4\end{bmatrix}$$

in the ordered basis $\mathcal{U}$ given by

$$\Big(\mathsf{u}_{1s} = \begin{bmatrix}1\\1\\1\end{bmatrix}, \mathsf{u}_{2s} = \begin{bmatrix}1\\2\\2\end{bmatrix}, \mathsf{u}_{3s} = \begin{bmatrix}1\\2\\3\end{bmatrix}\Big).$$

4.4.3.- Consider the vector space $V = \mathbb{P}_2$ with the ordered basis $\mathcal{S}$ given by

$$\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2\big).$$

(a) Find the components of the polynomial $\boldsymbol{r}(x) = 2 + 3x - x^2$ in the ordered basis $\mathcal{S}$.

(b) Find the components of the same polynomial $\boldsymbol{r}$ given in part (a) but now in the ordered basis $\mathcal{Q}$ given by

$$\big(\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = 1 - x,\ \boldsymbol{q}_2 = x + x^2,\big).$$

4.4.4.- Let $\mathcal{S}$ be the standard ordered basis of $\mathbb{R}^{2,2}$, that is,

$$\mathcal{S} = (\mathsf{E}_{11}, \mathsf{E}_{12}, \mathsf{E}_{21}, \mathsf{E}_{22}) \subset \mathbb{R}^{2,2},$$

with

$$\mathsf{E}_{11} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \quad \mathsf{E}_{12} = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix},$$

$$\mathsf{E}_{21} = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \quad \mathsf{E}_{22} = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}.$$

(a) Show that the ordered set $\mathcal{M}$ below is a basis of $\mathbb{R}^{2,2}$, where

$$\mathcal{M} = (\mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathsf{M}_4) \subset \mathbb{R}^{2,2},$$

with

$$\mathsf{M}_1 = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, \quad \mathsf{M}_2 = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix},$$

$$\mathsf{M}_3 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \quad \mathsf{M}_4 = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix},$$

where the matrices above are written in the standard basis.

(b) Consider the matrix A written in the standard basis $\mathcal{S}$,

$$\mathsf{A} = \begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}.$$

Find the components of the matrix A in the ordered basis $\mathcal{M}$.

Chapter 5. Linear transformations

5.1. Linear transformations

We now introduce the notion of a linear transformation between vector spaces. Vector spaces are defined not by the elements they are made of but by the relation among these elements. They are defined by the properties of the operation we called linear combinations of the vector space elements. Linear transformations are a very special type of functions that preserve the linear combinations on both vector spaces where they are defined. Linear transformations generalize the definition of a linear function introduced in Sect. 2.1 between the spaces $\mathbb{F}^n$ and $\mathbb{F}^m$ to any two vector spaces. Linear transformations are also called in the literature as linear maps or linear mappings.

Definition 5.1.1. *Given the vector spaces V and W over $\mathbb{F}$, the function $\boldsymbol{T} : V \to W$ is called a **linear transformation** iff for all $\boldsymbol{u}$, $\boldsymbol{v} \in V$ and all scalars a, $b \in \mathbb{F}$ holds*

$$\boldsymbol{T}(a\boldsymbol{u} + b\boldsymbol{v}) = a\,\boldsymbol{T}(\boldsymbol{u}) + b\,\boldsymbol{T}(\boldsymbol{v}).$$

*In the case that $V = W$ a linear transformation $\boldsymbol{T} : V \to V$ is called a **linear operator**.*

Taking $a = b = 0$ in the definition above we see that every linear transformation satisfies that $\boldsymbol{T}(\boldsymbol{0}) = \boldsymbol{0}$. As we said in Sect. 2.1, if a function does not satisfy this condition, then it cannot be a linear transformation. We now consider several examples of linear transformations. We do not prove that these examples are indeed linear transformations; the proofs are left to the reader.

Example 5.1.1: We give several examples of linear transformations.

(a) The transformation $\boldsymbol{I}_V : V \to V$ defined by $\boldsymbol{I}_V(\boldsymbol{v}) = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$ is called the *identity transformation*. The transformation $\boldsymbol{0} : V \to W$ defined by $\boldsymbol{0}(\boldsymbol{v}) = \boldsymbol{0}$ for all $\boldsymbol{v} \in V$ is called the *zero transformation*.

(b) Every example of a matrix as a linear function given in Sect. 2.1 is a linear transformation. More precisely, if $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, then any $m \times n$ matrix A defines a linear transformation $\boldsymbol{A} : \mathbb{F}^n \to \mathbb{F}^m$ by $\boldsymbol{A}(\mathsf{x}) = \mathsf{Ax}$. Therefore, rotations, reflections, projections, dilations are linear transformations in the sense given in Def. 5.1.1 above. In particular, square matrices are now called linear operators.

(c) The vector spaces V and W are part of the definition of the linear transformation. For example, a matrix A alone does not determine a linear transformation, since the vector spaces must also be specified. For example, and $m \times n$ matrix A defines two different linear transformations, the first one $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^m$, and the second one $\boldsymbol{A} : \mathbb{C}^n \to \mathbb{C}^m$. Although the action of these transformations is the same, the matrix-vector product $\boldsymbol{A}(\boldsymbol{x}) = \mathsf{Ax}$, the linear transformations are different. We will see in Chapter 9 that in the case $m = n$ these transformations may have different eigenvalues and eigenvectors.

(d) Let $V = \mathbb{P}_3$, $W = \mathbb{P}_2$ and let $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_2$ be the differentiation transformation

$$\boldsymbol{D}(\boldsymbol{p}) = \frac{d\boldsymbol{p}}{dx}.$$

That is, given a polynomial $\boldsymbol{p} \in \mathbb{P}_3$, the transformation $\boldsymbol{D}$ acting on $\boldsymbol{p}$ is a polynomial one degree less than $\boldsymbol{p}$ given by the derivative of $\boldsymbol{p}$. For example,

$$\boldsymbol{p}(x) = 2 + 3x + x^2 - x^3 \in \mathbb{P}_3 \quad \Rightarrow \quad \boldsymbol{D}(\boldsymbol{p})(x) = 3 + 2x - 3x^2 \in \mathbb{P}_2.$$

The transformation $\boldsymbol{D}$ is linear, since for all polynomials $\boldsymbol{p}$, $\boldsymbol{q} \in \mathbb{P}_3$ and scalars a, $b \in \mathbb{F}$ holds

$$\boldsymbol{D}(a\boldsymbol{p} + b\boldsymbol{q})(x) = \frac{d}{dx}\big[a\,\boldsymbol{p}(x) + b\,\boldsymbol{q}(x)\big] = a\,\frac{d\boldsymbol{p}(x)}{dx} + b\,\frac{d\boldsymbol{q}(x)}{dx} = a\,\boldsymbol{D}(\boldsymbol{p})(x) + b\,\boldsymbol{D}(\boldsymbol{q})(x).$$

(e) Notice that the differentiation transformation introduced above can also be defined as a linear operator: Let $V = W = \mathbb{P}_3$, and introduce $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_3$ with the same action as above, that is, $\boldsymbol{D}(\boldsymbol{p}) = \dfrac{d\boldsymbol{p}}{dx}$. The vector spaces used to define the transformation are important. The transformation defined in this example $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_3$ is different for the one defined above $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_2$. Although the action is the same, since the action of both transformations is to take the derivative of their arguments, the transformations are different. We comment on these issues later on.

(f) Let $V = \mathbb{P}_2$, $W = \mathbb{P}_3$ and let $\boldsymbol{S} : \mathbb{P}_2 \to \mathbb{P}_3$ be the integral transformation

$$\boldsymbol{S}(\boldsymbol{p})(x) = \int_0^x \boldsymbol{p}(t)\, dt.$$

For example,

$$\boldsymbol{p}(x) = 2 + 3x + x^2 \in \mathbb{P}_2 \quad \Rightarrow \quad \boldsymbol{S}(\boldsymbol{p})(x) = 2x + \frac{3}{2}\, x^2 + \frac{1}{3}\, x^3 \in \mathbb{P}_3.$$

The transformation $\boldsymbol{S}$ is linear, since for all polynomials $\boldsymbol{p}$, $\boldsymbol{q} \in \mathbb{P}_2$ and scalars a, $b \in \mathbb{F}$ holds

$$\boldsymbol{S}(a\boldsymbol{p} + b\boldsymbol{q})(x) = \int_0^x \big[a\, \boldsymbol{p}(t) + b\, \boldsymbol{q}(t)\big]\, dt = a \int_0^x \boldsymbol{p}(t)\, dt + b \int_0^x \boldsymbol{q}(t)\, dt = a\, \boldsymbol{S}(\boldsymbol{p})(x) + b\, \boldsymbol{S}(\boldsymbol{q})(x).$$

(g) Let $V = C^1\big(\mathbb{R}, \mathbb{R}\big)$, the space of all functions $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}$ having one continuous derivative,that is $\boldsymbol{f}'$ is continuous. Let $W = C^0\big(\mathbb{R}, \mathbb{R}\big)$, the space of all continuous functions $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}$. Notice that $V \subsetneq W$, since $\boldsymbol{f}(x) = |x|$, the absolute value function, belongs to W but it does not belong to V. The differentiation transformation $\boldsymbol{D} : V \to W$,

$$\boldsymbol{D}(\boldsymbol{f})(x) = \frac{d\boldsymbol{f}}{dx}(x),$$

is a linear transformation. The integral transformation $\boldsymbol{S} : W \to V$,

$$\boldsymbol{S}(\boldsymbol{f})(x) = \int_0^x \boldsymbol{f}(t)\, dt,$$

is also a linear transformation.

◁

5.1.1. **The null and range spaces.** The range and null spaces of an $m \times n$ matrix introduced in Sect. 2.5 can be also be defined for linear transformations between arbitrary vector spaces.

Definition 5.1.2. *Let V, W be vector spaces and $\boldsymbol{T} : V \to W$ be a linear transformation. The **null space** of the transformation $\boldsymbol{T}$ is the set $N(\boldsymbol{T}) \subset V$ given by*

$$N(\boldsymbol{T}) = \big\{ \boldsymbol{x} \in V \ : \ \boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{0} \ \big\}$$

*The **range space** of the transformation $\boldsymbol{T}$ is the set $R(\boldsymbol{T}) \subset W$ given by*

$$R(\boldsymbol{T}) = \big\{ \boldsymbol{y} \in W \ : \ \boldsymbol{y} = \boldsymbol{T}(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in V \big\}.$$

The null space of a linear transformation is also called the *kernel* of the transformation and denoted as $\ker(\boldsymbol{T})$. The range space of a linear transformation is also called the *image* of the transformation.

EXAMPLE 5.1.2: In the case of $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ and $\boldsymbol{A} = \mathsf{A}$ an $m \times n$ matrix, we have seen many examples of the sets $N(\mathsf{A})$ and $R(\mathsf{A})$ in Sect. 2.5. Additional examples in the case of more general linear transformations are the following:

(a) Let $V = \mathbb{P}_3$, $W = \mathbb{P}_2$, and let $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_2$ be the differentiation transformation

$$\boldsymbol{D}(\boldsymbol{p}) = \frac{d\boldsymbol{p}}{dx}.$$

Recalling that $\boldsymbol{D}(\boldsymbol{p}) = 0$ iff $\boldsymbol{p}(x) = c$, with c constant, we conclude that $N(\boldsymbol{D})$ is the set of constant polynomials, that is

$$N(\boldsymbol{D}) = \text{Span}\big(\{\boldsymbol{p}_0 = 1\}\big) \subset \mathbb{P}_3.$$

Let us now find the set $R(\boldsymbol{D})$. Notice that the derivative of a degree three polynomial is a degree two polynomial, never a degree three. We conclude that $R(\boldsymbol{D}) = \mathbb{P}_2$.

(b) Let $V = \mathbb{P}_2$, $W = \mathbb{P}_3$, and let $\boldsymbol{S} : \mathbb{P}_2 \to \mathbb{P}_3$ be the integration transformation

$$\boldsymbol{S}(\boldsymbol{p})(x) = \int_0^x \boldsymbol{p}(t)\, dt.$$

Recalling that $\boldsymbol{S}(\boldsymbol{p}) = 0$ iff $\boldsymbol{p}(x) = 0$, we conclude that

$$N(\boldsymbol{S}) = \big\{\boldsymbol{p}(x) = 0\big\} \subset \mathbb{P}_2.$$

Let us now find the set $R(\boldsymbol{S})$. Notice that the integral of a nonzero constant polynomial is a polynomial of degree one. We conclude that

$$R(\boldsymbol{S}) = \big\{\boldsymbol{p}(x) \in \mathbb{P}_3 \ : \ \boldsymbol{p}(x) = a_1 x + a_2 x^2 + a_3 x^3\big\}.$$

Therefore, non-zero constant polynomials do not belong to $R(\boldsymbol{S})$, so $R(\boldsymbol{S}) \subsetneq \mathbb{P}_3$.

◁

Given a linear transformation $\boldsymbol{T} : V \to W$ between vector spaces V and W, it is not difficult to show that the sets $N(\boldsymbol{T}) \subset V$ and $R(\boldsymbol{T}) \subset W$ are also subspaces of their respective vector spaces.

Theorem 5.1.3. *Let V and W be vector spaces and $\boldsymbol{T} : V \to W$ be a linear transformation. Then, the sets $N(\boldsymbol{T}) \subset V$ and $R(\boldsymbol{T}) \subset W$ are subspaces of V and W, respectively.*

The proof is formally the same as the proof of Theorem 2.5.3 in Sect. 2.5.

Proof of Theorem 5.1.3: The sets $N(\boldsymbol{T})$ and $R(\boldsymbol{T})$ are subspaces because the transformation $\boldsymbol{T}$ is linear. Consider two arbitrary elements $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in N(\boldsymbol{T})$, that is, $\boldsymbol{T}(\boldsymbol{x}_1) = \boldsymbol{0}$ and $\boldsymbol{T}(\boldsymbol{x}_2) = \boldsymbol{0}$. Then, for any a, $b \in \mathbb{F}$ holds

$$\boldsymbol{T}(a\boldsymbol{x}_1 + b\boldsymbol{x}_2) = a\, \boldsymbol{T}(\boldsymbol{x}_1) + b\, \boldsymbol{T}(\boldsymbol{x}_2) = \boldsymbol{0} \quad \Rightarrow \quad (a\boldsymbol{x}_1 + b\boldsymbol{x}_2) \in N(\boldsymbol{T}).$$

Therefore, $N(\boldsymbol{T}) \subset V$ is a subspace. Analogously, consider two arbitrary elements $\boldsymbol{y}_1$, $\boldsymbol{y}_2 \in R(\boldsymbol{T})$, that is, there exist $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in V$ such that $\boldsymbol{y}_1 = \boldsymbol{T}(\boldsymbol{x}_1)$ and $\boldsymbol{y}_2 = \boldsymbol{T}(\boldsymbol{x}_2)$. Then, for any a, $b \in \mathbb{F}$ holds

$$(a\boldsymbol{y}_1 + b\boldsymbol{y}_2) = a\, \boldsymbol{T}(\boldsymbol{x}_1) + b\, \boldsymbol{T}(\boldsymbol{x}_2) = \boldsymbol{T}(a\boldsymbol{x}_1 + b\boldsymbol{x}_2) \quad \Rightarrow \quad (a\boldsymbol{y}_2 + b\boldsymbol{y}_2) \in R(\boldsymbol{T}).$$

Therefore, $R(\boldsymbol{T}) \subset W$ is a subspace. This establishes the Theorem. □

5.1.2. **Injections, surjections and bijections.** We now specify useful properties a linear transformation might have.

Definition 5.1.4. *Let V and W be vector spaces and $\boldsymbol{T} : V \to W$ be a linear transformation.*

(a) $\boldsymbol{T}$ *is* ***injective*** *(or one-to-one) iff for all vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in V$ holds*

$$\boldsymbol{v}_1 \neq \boldsymbol{v}_2 \quad \Rightarrow \quad \boldsymbol{T}(\boldsymbol{v}_1) \neq \boldsymbol{T}(\boldsymbol{v}_2).$$

(b) $\boldsymbol{T}$ *is* ***surjective*** *(or onto) iff $R(\boldsymbol{T}) = W$.*

(c) $\boldsymbol{T}$ *is* ***bijective*** *(or an isomorphism) iff $\boldsymbol{T}$ is both injective and surjective.*

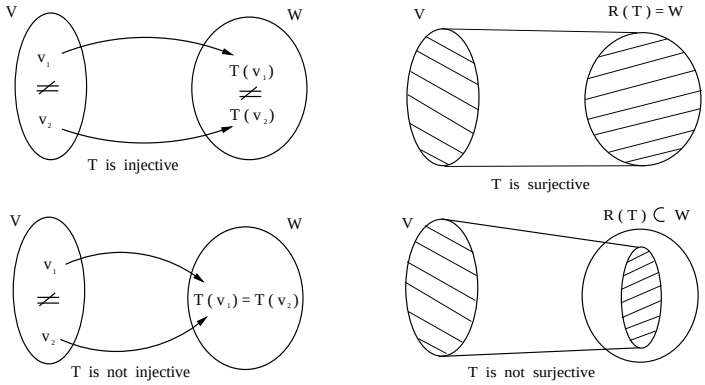

FIGURE 39. Sketch representing an injective and a non-injective function, as well as a surjective and a non-surjective function.

In Fig. 39 we sketch the meaning of these definitions using standard pictures from set theory. Notice that a given transformation can be only injective, or it can be only surjective, or it can be both injective and surjective (bijective), or it can be neither. That a transformation is injective does not imply anything about whether it is surjective or not. That a transformation is surjective does not imply anything about whether it is injective or not. Before we present examples of injective and/or surjective linear transformations it is useful to introduce a result to characterize those transformations that are injective. This can be done in terms of the null space of the transformation.

Theorem 5.1.5. *The linear transformation* $\boldsymbol{T}: V \to W$ *is injective iff* $N(\boldsymbol{T}) = \{\boldsymbol{0}\}$.

Proof of Theorem 5.1.5:
($\Rightarrow$) Since the transformation $\boldsymbol{T}$ is injective, given two elements $\boldsymbol{v} \neq \boldsymbol{0}$, we know that $\boldsymbol{T}(\boldsymbol{v}) \neq \boldsymbol{T}(\boldsymbol{0}) = \boldsymbol{0}$, where the last equality comes from the fact that $\boldsymbol{T}$ is linear. We conclude that the null space of $\boldsymbol{T}$ contains only the zero vector.
($\Leftarrow$) Since $N(\boldsymbol{T}) = \{\boldsymbol{0}\}$, given any two different elements $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, that is, $\boldsymbol{v}_1 - \boldsymbol{v}_2 \neq \boldsymbol{0}$, we know that $\boldsymbol{T}(\boldsymbol{v}_1 - \boldsymbol{v}_2) \neq \boldsymbol{0}$, because the only element in $N(\boldsymbol{T})$ is the zero vector. Since $\boldsymbol{T}(\boldsymbol{v}_1 - \boldsymbol{v}_2) = \boldsymbol{T}(\boldsymbol{v}_1) - \boldsymbol{T}(\boldsymbol{v}_2)$, we obtain that $\boldsymbol{T}(\boldsymbol{v}_1) \neq \boldsymbol{T}(\boldsymbol{v}_2)$. We conclude that $\boldsymbol{T}$ is injective. This establishes the Theorem. □

EXAMPLE 5.1.3: Let $V = \mathbb{R}^3$, $W = \mathbb{R}^2$, and consider the linear transformation $\boldsymbol{A}$ defined by the 2×3 matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}$. Is $\boldsymbol{A}$ injective? Is $\boldsymbol{A}$ surjective?

SOLUTION: A simple way to answer these questions is to find bases for the $N(\mathsf{A})$ and $R(\mathsf{A})$ spaces. Both bases can be obtained from the information given in $\mathsf{E_A}$, the reduced echelon form of A. A simple calculation shows

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathsf{E_A}.$$

This information gives a basis for $N(\mathsf{A})$, since all solutions $\mathsf{Ax} = \mathsf{0}$ are given by

$$\left.\begin{array}{l} x_1 = -2x_2, \\ x_2 \text{ free variable}, \\ x_3 = 0, \end{array}\right\} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} x_2 \quad \Rightarrow \quad N(\mathsf{A}) = \text{Span}\left(\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}\right).$$

Since $N(\mathsf{A}) \neq \{0\}$, the transformation associated with A is not injective. The reduced echelon form above also says that the first and third column vectors in A form a linearly independent set. Therefore, the set

$$\mathcal{R} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$$

is a basis for $R(\mathsf{A})$, so $\dim R(\mathsf{A}) = 2$, and then $R(\mathsf{A}) = \mathbb{R}^2$. Therefore, the transformation associated with A is surjective. ◁

EXAMPLE 5.1.4: Consider the differentiation and integration transformations

$$\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_2 \quad \boldsymbol{D}(\boldsymbol{p})(x) = \frac{d\boldsymbol{p}}{dx}(x), \qquad \boldsymbol{S} : \mathbb{P}_2 \to \mathbb{P}_3 \quad \boldsymbol{S}(\boldsymbol{p})(x) = \int_0^x \boldsymbol{p}(t)\, dt.$$

Show that $\boldsymbol{D}$ above is surjective but not injective, and $\boldsymbol{S}$ above is injective but not surjective.

SOLUTION: Let us start with the differentiation transformation. In Example 5.1.2 we found that $N(\boldsymbol{D}) \neq \{\boldsymbol{0}\}$ and $R(\boldsymbol{D}) = \mathbb{P}_2$. Therefore, $\boldsymbol{D}$ above is not injective but it is surjective. Regarding the integration transformation, we have seen in the same Example 5.1.2 that $N(\boldsymbol{S}) = \{\boldsymbol{0}\}$ and $R(\boldsymbol{S}) \subsetneq \mathbb{P}_3$. Therefore, $\boldsymbol{S}$ above is injective but it is not surjective. ◁

5.1.3. **Nullity-Rank Theorem.** The following result relates the dimensions of the null and range spaces of a linear transformation on finite dimensional vector spaces. This result is usually called Nullity-Rank Theorem, where the *nullity* of a linear transformation is the dimension of its null space, and the rank is the dimension of its range space. This result is also called the Dimension Theorem.

Theorem 5.1.6. *Every linear transformation $\boldsymbol{T} : V \to W$ between finite dimensional vector spaces V and W satisfies that*

$$\dim N(\boldsymbol{T}) + \dim R(\boldsymbol{T}) = \dim V. \tag{5.1}$$

Proof of Theorem 5.1.6: Let us denote by $n = \dim V$ and by $\mathcal{N} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k\}$ a basis for $N(\boldsymbol{T})$, where $0 \leqslant k \leqslant n$, with $k = 0$ representing the case $N(\boldsymbol{T}) = \{\boldsymbol{0}\}$. By Theorem 4.3.7 we know we can increase the set $\mathcal{N}$ into a basis of V, so let us denote this basis as

$$\mathcal{V} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_l\}, \qquad k + l = n.$$

In order to prove Eq. (5.1) we now show that the set

$$\mathcal{R} = \{\boldsymbol{T}(\boldsymbol{v}_1), \cdots, \boldsymbol{T}(\boldsymbol{v}_l)\}$$

is a basis of $R(\boldsymbol{T})$. We first show that $\text{Span}(\mathcal{R}) = R(\boldsymbol{T})$: Given any vector $\boldsymbol{v} \in V$, we can express it in the basis $\mathcal{V}$ as follows,

$$\boldsymbol{v} = x_1 \boldsymbol{u}_1 + \cdots + x_k \boldsymbol{u}_k + y_1 \boldsymbol{v}_1 + \cdots + y_l \boldsymbol{v}_l.$$

Since $R(\boldsymbol{T})$ is the set of vectors of the form $\boldsymbol{T}(\boldsymbol{v})$ for all $\boldsymbol{v} \in V$, therefore $\boldsymbol{T}(\boldsymbol{v}) \in R(\boldsymbol{T})$ iff

$$\boldsymbol{T}(\boldsymbol{v}) = y_1\, \boldsymbol{T}(\boldsymbol{v}_1) + \cdots + y_l\, \boldsymbol{T}(\boldsymbol{v}_l) \in \text{Span}(\mathcal{R}).$$

But this just says that $R(\boldsymbol{T}) = \text{Span}(\mathcal{R})$. Now we show that $\mathcal{R}$ is linearly independent: Given any linear combination of the form

$$\boldsymbol{0} = c_1\, \boldsymbol{T}(\boldsymbol{v}_1) + \cdots + c_l\, \boldsymbol{T}(\boldsymbol{v}_l) = \boldsymbol{T}(c_1 \boldsymbol{v}_1 + \cdots + c_l \boldsymbol{v}_l),$$

we conclude that $c_1 \boldsymbol{v}_1 + \cdots + c_l \boldsymbol{v}_l \in N(\boldsymbol{T})$, so there exist $d_1, \cdots, d_k$ such that

$$c_1 \boldsymbol{v}_1 + \cdots + c_l \boldsymbol{v}_l = d_1 \boldsymbol{u}_1 + \cdots + d_k \boldsymbol{u}_k,$$

which is equivalent to

$$d_1 \boldsymbol{u}_1 + \cdots + d_k \boldsymbol{u}_k - c_1 \boldsymbol{v}_1 - \cdots - c_l \boldsymbol{v}_l = \boldsymbol{0}.$$

Since $\mathcal{V}$ is a basis we conclude that all coefficients c_i and d_j must vanish, for $i = 1, \cdots, l$ and $j_1, \cdots, k$. This shows that $\mathcal{R}$ is a linearly independent set. Then $\mathcal{R}$ is a basis of $R(\boldsymbol{T})$ and so,

$$\dim N(\boldsymbol{T}) = k, \qquad \dim R(\boldsymbol{T}) = l = n - k.$$

This establishes the Theorem. □

One simple consequence of the Dimension Theorem is that an injective linear operator is also surjective, and vice versa.

Corollary 5.1.7. *A linear operator on a finite dimensional vector space is injective iff it is surjective.*

The proof is left as an exercise, Exercise **5.1.7**. In the case that the linear transformation is given by an $m \times n$ matrix, Theorem 5.1.6 establishes a relation between the the nullity and the rank of the matrix.

Corollary 5.1.8. *Every $m \times n$ matrix A satisfies that $\dim N(\mathsf{A}) + \operatorname{rank}(\mathsf{A}) = n$.*

Proof of Corollary 5.1.8: This Corollary is just a particular case of Theorem 5.1.6. Indeed, an $m \times n$ matrix A defines a linear transformation given by $\boldsymbol{A} : \mathbb{F}^n \to \mathbb{F}^m$ by $\boldsymbol{A}(\boldsymbol{x}) = \mathsf{A}\mathsf{x}$. Since $\operatorname{rank}(\mathsf{A}) = \dim R(\mathsf{A})$ and $\dim \mathbb{F}^n = n$, Eq. (5.1) implies that $\dim N(\mathsf{A}) + \operatorname{rank}(\mathsf{A}) = n$. This establishes the Corollary.

An alternative wording of this result uses E_A, the reduced echelon form of A. Since $N(\mathsf{A}) = N(\mathsf{E}_\mathsf{A})$, the $\dim N(\mathsf{A})$ is the number of non-pivot columns in E_A. We also know that $\operatorname{rank}(\mathsf{A})$ is the number of pivot columns in E_A. Therefore $\dim N(\mathsf{A}) + \operatorname{rank}(\mathsf{A})$ is the total number of columns in A, which is n. □

5.1.4. **Exercises.**

5.1.1.- Consider the operator $\boldsymbol{A} : \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix $\mathsf{A} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, which projects a vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ onto the line $x_1 = x_2$ on $\mathbb{R}^2$. Is $\boldsymbol{A}$ injective? Is $\boldsymbol{A}$ surjective?

5.1.2.- Fix any real number $\theta \in [0, 2\pi)$, and define the operator $\boldsymbol{R}(\theta) : \mathbb{R}^2 \to \mathbb{R}^2$ by the matrix $\mathsf{R}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$, which is a rotation by an angle θ counterclockwise. Is $\boldsymbol{R}(\theta)$ injective? Is $\boldsymbol{R}(\theta)$ surjective?

5.1.3.- Let $\mathbb{F}^{n,n}$ be the vector space of all $n \times n$ matrices, and fix $\mathsf{A} \in \mathbb{F}^{n,n}$. Determine which of the following transformations $\boldsymbol{T} : \mathbb{F}^{n,n} \to \mathbb{F}^{n,n}$ is linear.

(a) $\boldsymbol{T}(\mathsf{X}) = \mathsf{AX} - \mathsf{XA}$;
(b) $\boldsymbol{T}(\mathsf{X}) = \mathsf{X}^T$;
(c) $\boldsymbol{T}(\mathsf{X}) = \mathsf{X}^T + \mathsf{A}$;
(d) $\boldsymbol{T}(\mathsf{X}) = \mathsf{A}\,\mathrm{tr}\,(\mathsf{X})$.
(e) $\boldsymbol{T}(\mathsf{X}) = \mathsf{X} + \mathsf{X}^T$.

5.1.4.- Fix a vector $\mathsf{v} \in \mathbb{F}^n$ and then define the function $\boldsymbol{T} : \mathbb{F}^n \to \mathbb{F}$ by $\boldsymbol{T}(\mathsf{x}) = \mathsf{v}^T\mathsf{x}$. Show that $\boldsymbol{T}$ is a linear transformation. Is $\boldsymbol{T}$ a linear operator? Is $\boldsymbol{T}$ a linear functional?

5.1.5.- Show that the mapping $\Delta : \mathbb{P}_3 \to \mathbb{P}_1$ is a linear transformation, where

$$\Delta(\boldsymbol{p})(x) = \frac{d^2\boldsymbol{p}}{dx^2}(x)$$

Is Δ injective? Is Δ surjective?

5.1.6.- If the following statement is true, give a proof; if it is false, show it with an example. If a linear transformation on vector spaces $\boldsymbol{T} : V \to W$ is injective, then the image of a linearly independent set in V is a linearly independent set in W.

5.1.7.- Prove the following statement: A linear operator $\boldsymbol{T} : V \to V$ on a finite dimensional vector space V is injective iff $\boldsymbol{T}$ is surjective.

5.1.8.- Prove the following statement: If V and W are finite dimensional vector spaces with $\dim V > \dim W$, then every linear transformation $\boldsymbol{T} : V \to W$ is not injective.

5.2. Properties of linear transformations

5.2.1. **The inverse transformation.** Bijective transformations are invertible. We now introduce the inverse of a bijective transformation and we then show that if the invertible transformation is linear, then the inverse is again a linear transformation.

Definition 5.2.1. *The **inverse** of a bijective transformation $\boldsymbol{T} : V \to W$ is the transformation $\boldsymbol{T}^{-1} : W \to V$ defined for all $\boldsymbol{w} \in W$ and $\boldsymbol{v} \in V$ as follows,*

$$\boldsymbol{T}^{-1}(\boldsymbol{w}) = \boldsymbol{v} \quad \Leftrightarrow \quad \boldsymbol{T}(\boldsymbol{v}) = \boldsymbol{w}. \tag{5.2}$$

The inverse transformation defined above makes sense only in the case that the original transformation is bijective, hence bijective transformations are called *invertible*.

Example 5.2.1: Given a real number $\theta \in (0, \pi)$, define the transformation $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$

$$\boldsymbol{T}(\mathsf{x}) = \mathsf{R}_\theta\, \mathsf{x}, \qquad \mathsf{R}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Find the inverse transformation.

Solution: Since the transformation above is a rotation counterclockwise by an angle θ, the inverse transformation is a rotation clockwise by the same angle θ. In other words, $(\mathsf{R}_\theta)^{-1} = \mathsf{R}_{-\theta}$. Therefore, we conclude that,

$$\boldsymbol{T}^{-1} = (\mathsf{R}_\theta)^{-1} = \mathsf{R}_{-\theta} \quad \Rightarrow \quad (\mathsf{R}_\theta)^{-1} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix},$$

where we have used the relations $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$. ◁

It is common in the literature to define the inverse matrix in an equivalent way, which we state here as a Theorem.

Theorem 5.2.2. *The transformation $\boldsymbol{T}^{-1} : W \to V$ is the inverse of a bijective transformation $\boldsymbol{T} \in L(V)$ iff holds*

$$\boldsymbol{T}^{-1} \circ \boldsymbol{T} = \boldsymbol{I}_V, \qquad \boldsymbol{T} \circ \boldsymbol{T}^{-1} = \boldsymbol{I}_W.$$

Proof of Theorem 5.2.2:
($\Rightarrow$) Replace $\boldsymbol{w} = \boldsymbol{T}(\boldsymbol{v})$ into the expression of $\boldsymbol{T}^{-1}$, that is, $\boldsymbol{T}^{-1}(\boldsymbol{T}(\boldsymbol{v})) = \boldsymbol{v}$, for all $\boldsymbol{v} \in V$. This says that $\boldsymbol{T}^{-1} \circ \boldsymbol{T}$ is the identity operator in V. In a similar way, replace $\boldsymbol{v} = \boldsymbol{T}^{-1}(\boldsymbol{w})$ into the expression of $\boldsymbol{T}$, that is, $\boldsymbol{T}(\boldsymbol{T}^{-1}(\boldsymbol{w})) = \boldsymbol{w}$, for all $\boldsymbol{w} \in W$. This says that $\boldsymbol{T}^{-1} \circ \boldsymbol{T}$ is the identity in W.
($\Leftarrow$) If $\boldsymbol{w} = \boldsymbol{T}(\boldsymbol{v})$, then $\boldsymbol{T}^{-1}(\boldsymbol{w}) = \boldsymbol{T}^{-1}(\boldsymbol{T}(\boldsymbol{v})) = \boldsymbol{v}$. Conversely, if $\boldsymbol{v} = \boldsymbol{T}^{-1}(\boldsymbol{w})$, then $\boldsymbol{T}(\boldsymbol{v}) = \boldsymbol{T}(\boldsymbol{T}^{-1}(\boldsymbol{w})) = \boldsymbol{w}$. We conclude that

$$\boldsymbol{T}^{-1}(\boldsymbol{w}) = \boldsymbol{v} \quad \Leftrightarrow \quad \boldsymbol{T}(\boldsymbol{v}) = \boldsymbol{w}.$$

This establishes the Theorem. □

Using this theorem is simple to prove that the result in the previous example can be generalized to any linear transformation defined by an invertible matrix.

Example 5.2.2: Prove the following statement: If $\boldsymbol{T} : \mathbb{F}^n \to \mathbb{F}^n$ is given by $\boldsymbol{T}(\mathsf{x}) = \mathsf{A}\mathsf{x}$ with A invertible, then $\boldsymbol{T}$ is invertible and $\boldsymbol{T}^{-1}(\mathsf{y}) = A^{-1}\mathsf{y}$.

Solution: Denote $\boldsymbol{S}(\mathsf{y}) = A^{-1}\mathsf{y}$. Then $\boldsymbol{S}(\boldsymbol{T}(\mathsf{x})) = A^{-1}A\mathsf{x} = \mathsf{x}$ for all $\mathsf{x} \in \mathbb{F}^n$. This shows that $\boldsymbol{S} \circ \boldsymbol{T} = \boldsymbol{I}_n$. The transformation $\boldsymbol{S}$ also satisfies $\boldsymbol{T}(\boldsymbol{S}(\mathsf{y})) = A\,A^{-1}\mathsf{y} = \mathsf{y}$ for all $\mathsf{y} \in \mathbb{F}^n$. This shows that $\boldsymbol{T} \circ \boldsymbol{S} = \boldsymbol{I}_n$. Therefore, $\boldsymbol{T}^{-1} = \boldsymbol{S}$, that is, $\boldsymbol{T}^{-1}(\mathsf{y}) = A^{-1}\mathsf{y}$. ◁

EXAMPLE 5.2.3: Let $V = \mathbb{P}_3$, $W = \mathbb{P}_2$ and let $\boldsymbol{D} : V \to W$ be the differentiation transformation

$$\boldsymbol{D}(\boldsymbol{p}) = \frac{d\boldsymbol{p}}{dx}.$$

Is this transformation invertible?

SOLUTION: For every constant polynomial $\boldsymbol{p}(x) = c \in \mathbb{P}_3$, with $c \in \mathbb{F}$, holds $\boldsymbol{D}(c) = 0$. So the null space of the differentiation transformation is non-trivial, that is, $N(\boldsymbol{D}) \neq \{\boldsymbol{0}\}$. Since the transformation is not injective, it is not bijective. We conclude that the differentiation transformation above is not invertible. ◁

EXAMPLE 5.2.4: Let $V = \{\boldsymbol{p} \in \mathbb{P}_3 \ : \ \boldsymbol{p}(0) = 0\}$, $W = \mathbb{P}_2$ and let $\boldsymbol{D} : V \to W$ be the differentiation transformation

$$\boldsymbol{D}(\boldsymbol{p}) = \frac{d\boldsymbol{p}}{dx}.$$

Is this transformation invertible?

SOLUTION: First, notice that V is a vector space, since given any two elements $\boldsymbol{p}$, $\boldsymbol{q} \in V$, the linear combination satisfies $(a\boldsymbol{p}+b\boldsymbol{q})(0) = a\boldsymbol{p}(0)+b\boldsymbol{q}(0) = 0$ for all a, $b \in \mathbb{F}$, so $(a\boldsymbol{p}+b\boldsymbol{q}) \in V$. Also notice that the only constant polynomial $\boldsymbol{p}(x) = c$ belonging to V is the trivial one $c = 0$. So, the $N(\boldsymbol{D}) = \{\boldsymbol{0}\}$ and the differentiation operator is injective. It is simple to see that this operator is also surjective, since every element in W also belongs to $R(\boldsymbol{D})$. Indeed, for every constants a_2, a_1, $a_0 \in \mathbb{F}$ the arbitrary polynomial $\boldsymbol{r}(x) = a_2x^2+a_1x+a_0 \in W$ is the image under $\boldsymbol{D}$ of the polynomial $\boldsymbol{p}(x) = a_2x^3/3+a_1x^2/2+a_0x \in V$, that is, $\boldsymbol{D}(\boldsymbol{p}) = \boldsymbol{r}$. So, $W = R(\boldsymbol{D})$, the differentiation transformation is surjective, so it is bijective. We conclude that $\boldsymbol{D} : V \to W$ is invertible. ◁

We now show a couple of properties of the inverse transformation. We start showing that the inverse of a bijective linear transformation is also linear.

Theorem 5.2.3. *If the bijective transformation $T : V \to W$ is linear, then the inverse transformation $\boldsymbol{T}^{-1} : W \to V$ is also a linear transformation.*

Proof of Theorem 5.2.3: Since the linear transformation $\boldsymbol{T}$ is bijective, hence invertible, we now that for every pair of vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in V$ there exist vectors $\boldsymbol{w}_1$, $\boldsymbol{w}_2 \in W$ such that

$$\boldsymbol{T}(\boldsymbol{v}_1) = \boldsymbol{w}_1 \quad \Leftrightarrow \quad \boldsymbol{T}^{-1}(\boldsymbol{w}_1) = \boldsymbol{v}_1,$$
$$\boldsymbol{T}(\boldsymbol{v}_2) = \boldsymbol{w}_2 \quad \Leftrightarrow \quad \boldsymbol{T}^{-1}(\boldsymbol{w}_2) = \boldsymbol{v}_2.$$

Since $T : V \to W$ is linear, for all a_1, $a_2 \in \mathbb{F}$ holds

$$\begin{aligned} \boldsymbol{T}(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2) &= a_1\,\boldsymbol{T}(\boldsymbol{v}_1) + a_2\,\boldsymbol{T}(\boldsymbol{v}_2), \\ &= a_1\boldsymbol{w}_1 + a_2\boldsymbol{w}_2. \end{aligned}$$

By definition of the inverse transformation, the last equation above is equivalent to

$$\begin{aligned} \boldsymbol{T}^{-1}(a_1\boldsymbol{w}_1 + a_2\boldsymbol{w}_2) &= a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2, \\ &= a_1\,\boldsymbol{T}^{-1}(\boldsymbol{w}_1) + a_2\,\boldsymbol{T}^{-1}(\boldsymbol{w}_2). \end{aligned}$$

The last equation above says that $\boldsymbol{T}^{-1} : W \to V$ is a linear transformation. This establishes the Theorem. □

We now show that the inverse of a linear transformation is not only a linear transformation but it is also a bijection.

Theorem 5.2.4. *If the linear transformation $T : V \to W$ is a bijection, then the inverse linear transformation $\boldsymbol{T}^{-1} : W \to V$ also is a bijection.*

Proof of Theorem 5.2.4: Since $\boldsymbol{T}$ is a bijection, it is invertible so, for every $\boldsymbol{w}_1$, $\boldsymbol{w}_2 \in W$ there exist $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in V$ such that

$$\boldsymbol{T}^{-1}(\boldsymbol{w}_1) = \boldsymbol{v}_1 \quad \Leftrightarrow \quad \boldsymbol{T}(\boldsymbol{v}_1) = \boldsymbol{w}_1,$$
$$\boldsymbol{T}^{-1}(\boldsymbol{w}_2) = \boldsymbol{v}_2 \quad \Leftrightarrow \quad \boldsymbol{T}(\boldsymbol{v}_2) = \boldsymbol{w}_2.$$

In order to show that $\boldsymbol{T}^{-1}$ is injective pick up $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ with $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$ and show that this implies $\boldsymbol{v}_1 \neq \boldsymbol{v}_2$. This is indeed the case, since subtracting the first line from the second line above we get $\boldsymbol{0} \neq \boldsymbol{w}_2 - \boldsymbol{w}_1 = \boldsymbol{T}(\boldsymbol{v}_2 - \boldsymbol{v}_1)$. Since $\boldsymbol{T}$ is injective, its null space is trivial, so $\boldsymbol{v}_2 - \boldsymbol{v}_1 \neq \boldsymbol{0}$. We then conclude that $\boldsymbol{T}^{-1}$ is injective. This inverse transformation is also surjective, which is simple to see from the definition of inverse transformation: For all $\boldsymbol{v} \in V$ there exists $\boldsymbol{w} \in W$ such that

$$\boldsymbol{T}(\boldsymbol{v}) = \boldsymbol{w} \quad \Leftrightarrow \quad \boldsymbol{T}^{-1}(\boldsymbol{w}) = \boldsymbol{v}.$$

We conclude that $R(\boldsymbol{T}^{-1}) = V$, so $\boldsymbol{T}^{-1}$ is surjective. This establishes the Theorem. □

A simple corollary of Theorem 5.2.4 is that $\boldsymbol{T}^{-1}$ is invertible. It is not difficult to show that $\left(\boldsymbol{T}^{-1}\right)^{-1} = \boldsymbol{T}$. We have mentioned in Sect. 5.1 that a bijective linear transformation is also called and *isomorphism*. This is the reason for the following definition.

Definition 5.2.5. *The vector spaces V and W are called* ***isomorphic*** *iff there exist an isomorphism $T : V \to W$.*

Different vector spaces which are isomorphic are the same space from the point of view of linear algebra. This means that any linear combination performed in one of the spaces can be translated to the other space through the isomorphism, and any linear operator defined on one of these vector spaces can be translated into a unique linear operator on the other space. This translation is not unique, since there exist infinitely many isomorphisms between isomorphic vector spaces.

EXAMPLE 5.2.5: Show that the vector space $V = \{\boldsymbol{p} \in \mathbb{P}_3 \ : \ \boldsymbol{p}(0) = 0\}$ is isomorphic to $\mathbb{P}_2$.

SOLUTION: We have seen in Example 5.2.4 that the differentiation $\boldsymbol{D} : V \to \mathbb{P}_2$ is a bijection, that is, an isomorphism. Therefore, the spaces V and $\mathbb{P}_2$ are isomorphic. We mention that this isomorphism relates the polynomials in V and $\mathbb{P}_2$ as follows

$$V \ni \boldsymbol{p}(x) = a_3x^3 + a_2x^2 + a_1x \mapsto \frac{d\boldsymbol{p}}{dx} = 3a_3x^2 + 2a_2x + a_1 \in \mathbb{P}_2.$$

◁

EXAMPLE 5.2.6: Show whether the vector spaces $\mathbb{F}^{2,2}$ and $\mathbb{F}^4$ are isomorphic or not.

SOLUTION: In order to see if these vector spaces are isomorphic, we need to find a bijective transformation between the spaces. We propose the following transformation: $\boldsymbol{T} : \mathbb{F}^{2,2} \to \mathbb{F}^4$ given by

$$\boldsymbol{T}\left(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}\right) = \begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{bmatrix}.$$

It is simple to see that $\boldsymbol{T}$ is a bijection, so we conclude that $\mathbb{F}^{2,2}$ is isomorphic to $\mathbb{F}^4$. We mention that this is not the only isomorphism between these vector spaces. Another

isomorphism is the following:

$$\boldsymbol{S}\Big(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}\Big) = \begin{bmatrix} x_{22} \\ x_{21} \\ x_{12} \\ x_{11} \end{bmatrix}.$$

◁

Example 5.2.7: Show whether the vector spaces $\mathbb{P}_2(\mathbb{F},\mathbb{F})$ and $\mathbb{F}^3$ are isomorphic or not.

Solution: In order to see if these vector spaces are isomorphic, we need to find a bijective transformation between the spaces. We propose the following transformation: $\boldsymbol{T}:\mathbb{P}_2 \to \mathbb{F}^3$ given by

$$\boldsymbol{T}\big(a_2x^2 + a_1x + a_0\big) = \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix}.$$

It is simple to see that $\boldsymbol{T}$ is a bijection, so we conclude that $\mathbb{P}_2$ is isomorphic to $\mathbb{F}^3$. We mention that this is not the only isomorphism between these vector spaces. Another isomorphism is the following:

$$\boldsymbol{S}\big(a_2x^2 + a_1x + a_0\big) = \begin{bmatrix} a_0 \\ a_2 \\ a_1 \end{bmatrix}.$$

◁

Theorem 5.2.6. *Every n-dimensional vector space over the field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.*

The proof is left as Exercise **5.2.5**. This result says that nothing essentially different from $\mathbb{F}^n$ should be expected from any finite dimensional vector space. On the other hand, infinite dimensional vector spaces are essentially different from $\mathbb{F}^n$. Several results that hold for finite dimensional vector spaces do not hold for infinite dimensional ones. An example is the Nullity-Rank Theorem stated in Section 5.1. One could say that finite dimensional vector spaces are all alike; every infinite dimensional vector space is special in its own way.

5.2.2. **The vector space of linear transformations.** Linear transformations from V to W can be combined in different ways to obtain new linear transformations. Two linear transformations can be added together and a linear transformation can be multiplied by a scalar. The space of all linear transformations with these operations is a vector space. This is why a linear transformation can be seen either as a function between vector spaces or as a vector in the space of all linear transformations. As an example of these ideas recall the set of all $m \times n$ matrices, denoted as $\mathbb{F}^{m,n}$. This set is a vector space, since two matrices can be added together and a matrix can be multiplied by a scalar. Since any element in this set, an $m \times n$ matrix A, is also a linear transformation $\mathsf{A}:\mathbb{F}^n \to \mathbb{F}^m$, the set of all linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^m$ is a vector space. This is why an $m \times n$ matrix A can be interpreted either as a linear transformation between vector spaces $\mathsf{A}:\mathbb{F}^n \to \mathbb{F}^m$ or as a vector in the space of matrices, $\mathsf{A} \in \mathbb{F}^{m,n}$.

Definition 5.2.7. *Given the vector spaces V and W over $\mathbb{F}$, we denote by $L(V,W)$ the **set of all linear transformations** from V to W. The set $L(V,V)$ containing all linear operators from V to V is denoted as $L(V)$. Furthermore, given $\boldsymbol{T}$, $\boldsymbol{S} \in L(V,W)$ and scalars a, $b \in \mathbb{F}$, introduce the **linear combination** of linear transformations as follows*

$$(a\,\boldsymbol{T} + b\boldsymbol{S})(\boldsymbol{v}) = a\,\boldsymbol{T}(\boldsymbol{v}) + b\,\boldsymbol{S}(\boldsymbol{v}) \qquad \text{for all } \boldsymbol{v} \in V.$$

Notice that the zero transformation $\boldsymbol{0} : V \to W$, given by $\boldsymbol{0}(\boldsymbol{v}) = \boldsymbol{0}$ for all $\boldsymbol{v} \in V$, belongs to $L(V, W)$. Moreover, the space $L(V, W)$ with the addition and scalar multiplication operations above is indeed a vector space.

Theorem 5.2.8. *The set $L(V, W)$ of linear transformations from V to W together with the addition and scalar multiplication operations in Definition 5.2.7 is a vector space.*

The proof is to verify all properties in the Definition 4.1.1, and we left it to the reader. We conclude that a linear transformations can be interpreted either as linear maps between vector spaces or as vectors in the space $L(V, W)$. Furthermore, for finite dimensional vector spaces we have the following result.

Theorem 5.2.9. *If V and W are finite dimensional vector spaces, then $L(V, W)$ is also finite dimensional and $\dim L(V, W) = (\dim V)(\dim W)$.*

Proof of Theorem 5.2.9: We introduce a set of linear transformations and we show that this set is indeed a basis for $L(V, W)$. Before that let us denote $n = \dim V$, $m = \dim W$, and let us introduce both a basis $\mathcal{V} = \{\boldsymbol{v}_i\}$ of V and a basis $\mathcal{W} = \{\boldsymbol{w}_j\}$ of W, where the indices take values $i = 1, \cdots, n$ and $j = 1, \cdots, m$. The proposal for a basis of $L(V, W)$ is the set $\mathcal{S} = \{\boldsymbol{S}_{ij}\}$, with elements defined as follows,

$$\boldsymbol{S}_{ij}(\boldsymbol{v}_k) = \begin{cases} \boldsymbol{w}_j & \text{for } i = k, \\ \boldsymbol{0} & \text{for } i \neq k, \end{cases} \quad \text{where } i = 1, \cdots, n \text{ and } j = 1, \cdots, m.$$

So, the map $\boldsymbol{S}_{ij}$ transforms the basis vector $\boldsymbol{v}_i$ into the basis vector $\boldsymbol{w}_j$ and the rest of the basis vectors in $\mathcal{V}$ into the zero vector in W. We now show that $\mathcal{S}$ is a basis for $L(V, W)$. Let us first prove that $\mathcal{S}$ spans $L(V, W)$. Indeed, for every linear transformation $\boldsymbol{T} \in L(V, W)$ and every vector $\boldsymbol{v} \in V$ holds,

$$\boldsymbol{T}(\boldsymbol{v}) = \sum_{i=1}^{n} v_i \, \boldsymbol{T}(\boldsymbol{v}_i),$$

where we used the decomposition $\boldsymbol{v} = \sum_{i=1}^{n} v_i \boldsymbol{v}_i$ of vector $\boldsymbol{v}$ in the basis $\mathcal{V}$. Since $\boldsymbol{T}(\boldsymbol{v}_i) \in W$, it can be decomposed in the $\mathcal{W}$ basis as follows,

$$\boldsymbol{T}(\boldsymbol{v}_i) = \sum_{j=1}^{m} T_{ij} \, \boldsymbol{w}_j.$$

Replacing the last equation on the previous one we get,

$$\boldsymbol{T}(\boldsymbol{v}) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} v_i T_{ij} \, \boldsymbol{w}_j \Big),$$

The basis vector $\boldsymbol{w}_j$ can be written in terms of the maps $\boldsymbol{S}_{ij}$ as $\boldsymbol{w}_j = \boldsymbol{S}_{kj}(\boldsymbol{v}_k)$. Choosing index k to be the value of index i in the sums above we get,

$$\boldsymbol{T}(\boldsymbol{v}) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} v_i T_{ij} \, \boldsymbol{S}_{ij}(\boldsymbol{v}_i) \Big) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} T_{ij} \, \boldsymbol{S}_{ij}(v_i \boldsymbol{v}_i) \Big).$$

Here is the critical step: Since $\boldsymbol{S}_{ij}(\boldsymbol{v}_k) = \boldsymbol{0}$ for $k \neq i$, the following equation holds, $\boldsymbol{S}_{ij}(v_i \boldsymbol{v}_i) = \boldsymbol{S}_{ij}(\boldsymbol{v})$. Introducing this property in the equation above we get

$$\boldsymbol{T}(\boldsymbol{v}) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} T_{ij} \, \boldsymbol{S}_{ij}(\boldsymbol{v}) \Big) \quad \text{for all } \boldsymbol{v} \in V,$$

that is,

$$\boldsymbol{T} = \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}\,\boldsymbol{S}_{ij} \quad \Rightarrow \quad \boldsymbol{T} \in \mathrm{Span}\big(\mathcal{S}\big) \subset L(V,W).$$

Since the map $\boldsymbol{T}$ is an arbitrary element in $L(V,W)$, we conclude that $L(V,W) \subset \mathrm{Span}\big(\mathcal{S}\big)$, hence $\mathrm{Span}\big(\mathcal{S}\big) = L(V,W)$. We now show that $\mathcal{S}$ is a linearly independent set. Indeed, suppose there exist scalars c_{ij} satisfying

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\,\boldsymbol{S}_{ij} = \boldsymbol{0}.$$

Interchange the sums and evaluate this expression at the basis vector $\boldsymbol{v}_k$, we obtain,

$$\boldsymbol{0} = \sum_{j=1}^{m}\Big(\sum_{i=1}^{n} c_{ij}\,\boldsymbol{S}_{ij}(\boldsymbol{v}_k)\Big) = \sum_{j=1}^{m} c_{kj}\,\boldsymbol{w}_j.$$

Since the set $\mathcal{W}$ is a basis of W, this implies that the coefficients $c_{kj} = 0$ for $j = 1, \cdots, m$. The same argument holds for every $k = 1, \cdots, n$, so we conclude that every coefficient c_{ij} vanishes. The set $\mathcal{S}$ is linearly independent, and so, it is a basis for $L(V,W)$. Since the set contains nm elements, we conclude that $\dim L(V,W) = \big(\dim V\big)\big(\dim W\big)$. This establishes the Theorem. □

5.2.3. **Linear functionals and the dual space.** An important example of linear transformations is the case where the target vector space W is the set of scalars $\mathbb{F}$, the latter, let us recall, is always a vector space. We often use boldface Greek letters to denote linear functionals.

Definition 5.2.10. *Given the vector space V over $\mathbb{F}$, the linear transformation $\boldsymbol{\phi} : V \to \mathbb{F}$ is called a **linear functional**. The vector space of linear functionals $L(V,\mathbb{F})$ is denoted as V^* and called the **dual space of** V.*

EXAMPLE 5.2.8: Show that the projection onto the first component $\boldsymbol{\pi}_1 : \mathbb{F}^3 \to \mathbb{F}$ is a linear functional, where

$$\boldsymbol{\pi}_1\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\Big) = x_1.$$

SOLUTION: We only need to show that $\boldsymbol{\pi}_1$ is a linear transformation. This is indeed the case, since for all x, $\mathsf{y} \in \mathbb{F}^3$ and all a, $b \in \mathbb{F}$ holds

$$\boldsymbol{\pi}_1(a\mathsf{x} + b\mathsf{y}) = ax_1 + by_1 = a\,\boldsymbol{\pi}_1(\mathsf{x}) + b\,\boldsymbol{\pi}_1(\mathsf{y}).$$

Since the map is defined from $\mathbb{F}^3$ into $\mathbb{F}$, we conclude that $\boldsymbol{\pi}_1$ is a linear functional. ◁

EXAMPLE 5.2.9: Consider the vector space $V = C\big([a,b],\mathbb{R}\big)$ of continuous real-valued functions with domain in the interval $[a,b] \subset \mathbb{R}$. Show that the function $\boldsymbol{\phi} : V \to \mathbb{R}$ given below is a linear functional, where

$$\boldsymbol{\phi}(\boldsymbol{f}) = \int_a^b \boldsymbol{f}(x)\,dx.$$

SOLUTION: We only need to show that the function $\boldsymbol{\phi} : V \to \mathbb{R}$ is a linear transformations. This is the case, since for all $\boldsymbol{f}$, $\boldsymbol{g} \in V$ and all scalars c, $d \in \mathbb{R}$ holds,

$$\begin{aligned}\boldsymbol{\phi}(c\boldsymbol{f} + d\boldsymbol{g}) &= \int_a^b \big(c\,\boldsymbol{f}(x) + d\,\boldsymbol{g}(x)\big)\,dx \\ &= c\int_a^b \boldsymbol{f}(x)\,dx + d\int_a^b \boldsymbol{g}(x)\,dx \\ &= c\,\boldsymbol{\phi}(\boldsymbol{f}) + d\,\boldsymbol{\phi}(\boldsymbol{g}).\end{aligned}$$

We conclude that $\boldsymbol{\phi} \in V^*$. ◁

EXAMPLE 5.2.10: Show that the trace map $\text{tr} : \mathbb{F}^{n,n} \to \mathbb{F}$ is a linear functional.

SOLUTION: We have already seen that the trace map is a linear map. Since the trace of a matrix is a scalar, we conclude that the trace map is a linear functional, so $\text{tr} \in \big(\mathbb{F}^{n,n}\big)^*$. ◁

The following result says that the dual space of a finite dimensional vector space is isomorphic to the original vector space. This means that the vector space and its dual are the same from the point of view of linear algebra. This result is not true for infinite dimensional vector spaces. This is one reason why infinite dimensional vector spaces, like function spaces, have more structure and are more complicated to study than finite dimensional vector spaces.

Theorem 5.2.11. *Every finite dimensional vector space is isomorphic to its dual space.*

We give two proofs of the same result. In the first proof, we show that an n-dimensional vector space V is isomorphic to $\mathbb{F}^n$, and that V^* is isomorphic to $\mathbb{F}^n$. We conclude that V is isomorphic to V^*. The second proof is more abstract, and makes use of what is called a dual basis of V^*.

Proof of Theorem 5.2.11: (First proof.) We already know that the ordered basis $\mathcal{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ of an n-dimensional vector space V determines the isomorphism $[\,]_v : V \to \mathbb{F}^n$, called the coordinate map, as follows

$$[\boldsymbol{x}]_v = \mathsf{x}_v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_v \quad \Leftrightarrow \quad \boldsymbol{x} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n.$$

So, any vector $\boldsymbol{x} \in V$ can be associated with a column vector $\mathsf{x}_v \in \mathbb{F}^n$. We now show that the dual space V^* is also isomorphic to the vector space $\mathbb{F}^n$. Once this is shown we can conclude that V is isomorphic to V^*. An arbitrary linear functional $\boldsymbol{\phi} \in V^*$ satisfies that

$$\begin{aligned}\boldsymbol{\phi}(\boldsymbol{x}) &= \boldsymbol{\phi}(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n) \\ &= x_1\,\boldsymbol{\phi}(\boldsymbol{v}_1) + \cdots + x_n\,\boldsymbol{\phi}(\boldsymbol{v}_n) \\ &= \big[\boldsymbol{\phi}(\boldsymbol{v}_1), \cdots, \boldsymbol{\phi}(\boldsymbol{v}_n)\big]_v \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_v \in \mathbb{F}\end{aligned}$$

for all $\boldsymbol{x} \in V$. Denoting the scalars $\phi_i = \boldsymbol{\phi}(\boldsymbol{v}_i)$, for $i = 1, \cdots, n$, we introduce the coordinate map $[\,]_v : V^* \to \mathbb{F}^n$ as follows

$$[\boldsymbol{\phi}]_v = \big[\phi_1, \cdots, \phi_n\big]_v \quad \Leftrightarrow \quad \boldsymbol{\phi}(\boldsymbol{x}) = \big[\phi_1, \cdots, \phi_n\big]_v \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_v.$$

The coordinate map defined on the space V^* of linear functionals associates a linear functional with a row vector. We have chosen a row vector instead of a column vector so we can use the matrix-vector product to express the action of $\boldsymbol{\phi}$ on a vector $\boldsymbol{x}$. We conclude that, once a basis is fixed in a finite dimensional vector space, vectors can be associated with column vectors, while linear functionals can be associated with row vectors. This establishes the Theorem. □

Proof of Theorem 5.2.11: (Second proof.) Since $V^* = L(V, \mathbb{F})$, by Theorem 5.2.9 we know that $\dim V^* = \dim V$, where we used the fact the space of scalars, $\mathbb{F}$, is itself a vector space and $\dim \mathbb{F} = 1$. From the proof of Theorem 5.2.9 we know that a basis of V^* is given by the set of linear functionals $\Phi = \{\boldsymbol{\phi}_i\}$ defined as follows: Given a basis $\mathcal{V} = \{\boldsymbol{v}_i\}$ of V, for $i = 1 \cdots, n = \dim V$, the linear functionals $\boldsymbol{\phi}_i$ satisfy

$$\boldsymbol{\phi}_i : V \to \mathbb{F}, \quad \boldsymbol{\phi}_i(\boldsymbol{v}_k) = \begin{cases} 1 & \text{for } i = k, \\ 0 & \text{for } i \neq k, \end{cases} \quad i, k = 1, \cdots, n.$$

Such a basis Φ is called the dual basis of $\mathcal{V}$. Now that we have a basis in V and a basis in V^* it is simple to find an isomorphism between these spaces. We propose the map $\boldsymbol{R} : V \to V^*$ that changes one basis into the other,

$$\boldsymbol{R}(\boldsymbol{v}_i) = \boldsymbol{\phi}_i.$$

We now show that this linear transformation $\boldsymbol{R}$ is an isomorphism. It is injective because of the following argument. Consider an arbitrary element $\boldsymbol{v} \in N(\boldsymbol{R})$, that is,

$$\boldsymbol{R}(\boldsymbol{v}) = \boldsymbol{0},$$

where $\boldsymbol{0} : V \to \mathbb{F}$ is the zero map. Introducing the basis decomposition $\boldsymbol{v} = \sum_{i=1}^n v_i \, \boldsymbol{v}_i$ in the expression above we get

$$\boldsymbol{0} = \sum_{i=1}^n v_i \, \boldsymbol{R}(\boldsymbol{v}_i) = \sum_{i=1}^n v_i \, \boldsymbol{\phi}_i.$$

Since the set Φ is a basis of V^*, then Φ is linearly independent, which implies that all coefficients v_i above vanish. We conclude that the vector $\boldsymbol{v} = \sum_{i=1}^n v_i \, \boldsymbol{v}_i = \boldsymbol{0}$, that is, the null space of $\boldsymbol{R}$ is trivial. Hence $\boldsymbol{R}$ is injective. This property together with the Nullity-Rank Theorem and the fact that $\dim V^* = \dim V$ imply that $\boldsymbol{R}$ is also surjective. We conclude that V is isomorphic to V^*. This establishes the Theorem. □

In the second proof above we introduced a particular basis of the dual space, called the dual basis.

Definition 5.2.12. *Given an n-dimensional vector space V with a basis $\mathcal{V} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$, the **dual basis of $\mathcal{V}$** is a particular basis $\Phi = \{\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_n\}$ of the dual space V^* that for $i, j = 1, \cdots, n$ the basis linear functionals $\boldsymbol{\phi}_i$ satisfy*

$$\boldsymbol{\phi}_i(\boldsymbol{v}_j) = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

EXAMPLE 5.2.11: Find the dual basis of the basis $\mathcal{U} \subset \mathbb{R}^2$, where

$$\mathcal{U} = \left\{ \mathsf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

SOLUTION: We need to find a set $\Phi = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}$ of linear functionals satisfying the equations

$$\boldsymbol{\phi}_1(\mathsf{u}_1) = 1, \qquad \boldsymbol{\phi}_2(\mathsf{u}_1) = 0,$$
$$\boldsymbol{\phi}_1(\mathsf{u}_2) = 0, \qquad \boldsymbol{\phi}_2(\mathsf{u}_2) = 1.$$

The two equation on the left can be solved independently of the two equations on the right. Recall that the most general expression for the linear functionals $\phi_1 : \mathbb{R}^2 \to \mathbb{R}$ and $\phi_2 : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\phi_1\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = a_1 x_1 + a_2 x_2. \qquad \phi_2\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = b_1 x_1 + b_2 x_2.$$

We only need to find the components a_1, a_2 and b_1, b_2. The a's can be obtained from the two equations on the left above, that is,

$$\left.\begin{array}{lll} \phi_1(\mathsf{u}_1) = 1 & \Rightarrow & \phi_1\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big) = a_1 + a_2 = 1 \\ \phi_1(\mathsf{u}_2) = 0 & \Rightarrow & \phi_1\Big(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\Big) = -a_1 + a_2 = 0. \end{array}\right\} \quad \Rightarrow \quad \begin{array}{l} a_1 = 1/2, \\ a_2 = 1/2. \end{array}$$

A similar calculation for ϕ_2 implies,

$$\left.\begin{array}{lll} \phi_2(\mathsf{u}_1) = 0 & \Rightarrow & \phi_2\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big) = b_1 + b_2 = 0 \\ \phi_2(\mathsf{u}_2) = 1 & \Rightarrow & \phi_1\Big(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\Big) = -b_1 + b_2 = 1. \end{array}\right\} \quad \Rightarrow \quad \begin{array}{l} b_1 = -1/2, \\ b_2 = 1/2. \end{array}$$

So we conclude that the dual basis of $\mathcal{U}$ is the basis $\Phi = \{\phi_1, \phi_2\}$ defined as follows

$$\phi_1\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \frac{1}{2}(x_1 + x_2), \qquad \phi_1\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \frac{1}{2}(-x_1 + x_2).$$

If we associate the linear functionals above with row vectors in $\mathbb{R}^2$ with the isomorphism

$$[\phi\,] = \begin{bmatrix} a_1, a_2 \end{bmatrix} \quad \Leftrightarrow \quad \phi\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = a_1 x_1 + a_2 x_2,$$

then, the answer of the Example is given by the row vectors

$$[\phi_1] = \frac{1}{2}\begin{bmatrix} 1, 1 \end{bmatrix}, \qquad [\phi_2] = \frac{1}{2}\begin{bmatrix} -1, 1 \end{bmatrix}.$$

Associating linear functionals components with row vector is useful, because the matrix-vector product can be used to represent the action of a functional onto a vector. The scalar obtained by the action of a linear functional onto a vector is the scalar given by the matrix-vector product of a row vector times a column vector. For example,

$$\phi_1(\mathsf{u}_1) = \frac{1}{2}\begin{bmatrix} 1, 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1, \qquad \phi_1(\mathsf{u}_2) = \frac{1}{2}\begin{bmatrix} 1, 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = 0,$$

$$\phi_2(\mathsf{u}_1) = \frac{1}{2}\begin{bmatrix} -1, 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0, \qquad \phi_2(\mathsf{u}_2) = \frac{1}{2}\begin{bmatrix} -1, 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = 1.$$

◁

5.2.4. **Exercises.**

5.2.1.- Show that the linear transformation $\boldsymbol{T} : \mathbb{F}^2 \to \mathbb{F}^2$ given below is invertible and find the inverse transformation, where

$$\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} x_1 + x_2 \\ 3x_1 - 2x_2 \end{bmatrix}.$$

5.2.2.- Consider the vector space

$$V = \{\boldsymbol{p} \in \mathbb{P}_4 \ : \ \boldsymbol{p}(0) = 0, \ \frac{d\boldsymbol{p}}{\delta x}(0) = 0\}.$$

Then show that the linear transformation $\Delta : V \to \mathbb{P}_2$ given below is invertible and find the inverse transformation, where

$$\Delta(\boldsymbol{p}) = \frac{d^2\boldsymbol{p}}{dx^2}.$$

5.2.3.- Use the Nullity-Rank Theorem 5.1.6 to show that if the finite dimensional vector spaces V and W are isomorphic, then they have the same dimension.

5.2.4.- Show that the spaces $\mathbb{P}_n$ and $\mathbb{F}^{n+1}$ are isomorphic.

5.2.5.- Use the coordinate map to prove Theorem 5.2.6: Every n-dimensional vector space over the field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.

5.2.6.- Show that the projection onto the i-component $\boldsymbol{\pi}_I : \mathbb{F}^n \to \mathbb{F}$ is a linear functional, where

$$\boldsymbol{\pi}_i\Big(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\Big) = x_i.$$

5.2.7.- Let $V = \mathbb{P}_2([0,1])$ be the vector space of polynomials up to degree two on the domain $[0,1] \subset \mathbb{R}$. Show that the function $\boldsymbol{\phi} : V \to \mathbb{R}$ given below is a linear functional, where

$$\boldsymbol{\phi}(\boldsymbol{p}) = \int_0^1 \boldsymbol{p}(x)\, dx.$$

5.2.8.- Given the basis $\mathcal{U} \subset \mathbb{R}^2$ below, find its dual basis, where

$$\mathcal{U} = \Big\{\mathsf{u}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}\Big\}.$$

5.3. The algebra of linear operators

A vector space where one can multiply two vectors to obtain a third vector is called an algebra. We have seen that the set $L(V,W)$ is a vector space. A multiplication can be defined between linear transformations in many different ways, but the most interesting one from the physical point of view is the composition of linear transformations. If $\boldsymbol{T}: U \to V$ and $\boldsymbol{S}: V \to W$ are linear transformations, then $\boldsymbol{S} \circ \boldsymbol{T}: U \to W$ defined as

$$(\boldsymbol{S} \circ \boldsymbol{T})(\boldsymbol{u}) = \boldsymbol{S}\big(\boldsymbol{T}(\boldsymbol{u})\big) \qquad \text{for all } \boldsymbol{u} \in U,$$

is also a linear transformation. This operation is not defined on a single vector space, but on three different spaces, namely, $L(U,V)$, $L(V,W)$ and $L(U,W)$. However, in the particular case that $U = V = W$ the composition of linear operators is defined on a single vector space $L(V,V)$, denoted simply as $L(V)$. The vector space $L(V)$ is an algebra of linear operators. We start introducing the definition of an algebra over a field.

Definition 5.3.1. *An* ***algebra*** V *over the field* $\mathbb{F}$ *is a vector space* V *over the field* $\mathbb{F}$ *equipped with an operation* $V \times V \to V$ *called* ***multiplication****, satisfying,*

$$\boldsymbol{u}(a\boldsymbol{v} + b\boldsymbol{w}) = a\,\boldsymbol{u}\boldsymbol{v} + b\,\boldsymbol{u}\boldsymbol{w} \qquad \text{for all } \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V \text{ and } a, b \in \mathbb{F}.$$

The algebra is called ***associative*** *iff the product satisfies*

$$\boldsymbol{u}(\boldsymbol{v}\boldsymbol{w}) = (\boldsymbol{u}\boldsymbol{v})\boldsymbol{w} \qquad \text{for all } \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V.$$

The algebra is called ***commutative*** *iff the product satisfies*

$$\boldsymbol{u}\boldsymbol{v} = \boldsymbol{v}\boldsymbol{u} \qquad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in V.$$

An algebra is called with ***identity*** *iff there is an element* $\boldsymbol{i} \in V$ *satisfying*

$$\boldsymbol{i}\boldsymbol{u} = \boldsymbol{u}\boldsymbol{i} = \boldsymbol{u} \qquad \text{for all } \boldsymbol{u} \in V.$$

Maybe the best known example of an algebra is the vector space $\mathbb{R}^3$ equipped with the cross product, also called vector product. This is a non-associative, non-commutative algebra.

Example 5.3.1: Let $\mathcal{S} = (\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3)$ be the standard ordered basis of $\mathbb{R}^3$ and introduce the cross product of vectors $\mathsf{u}, \mathsf{v} \in \mathbb{R}^3$ as follows

$$\mathsf{u} \times \mathsf{v} = (u_2 v_3 - u_3 v_2)\,\mathsf{e}_1 - (u_1 v_3 - u_3 v_1)\,\mathsf{e}_2 + (u_1 v_2 - u_2 v_1)\,\mathsf{e}_3,$$

where $\mathsf{u} = u_1\,\mathsf{e}_1 + u_2\,\mathsf{e}_2 + u_3\,\mathsf{e}_3$ and $\mathsf{v} = v_1\,\mathsf{e}_1 + v_2\,\mathsf{e}_2 + v_3\,\mathsf{e}_3$. Show that $\mathbb{R}^3$ equipped with the cross product is a non-associative, non-commutative algebra.

Solution: It is convenient to express the cross product of two vectors using the determinant notation. Indeed, the cross product above is equal to the following expression

$$\mathsf{u} \times \mathsf{v} = \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},$$

where in the first row of the array above contains the basis vectors and the other two rows contain vector components. So, the array above is not a matrix, but the definition of determinant of a matrix when used on this array produces the cross product of two vectors. This notation is convenient since the cross product shares all the properties that the determinant has. For example, the cross product satisfies the following two properties:

$$\mathsf{u} \times (a\mathsf{v}) = \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ a v_1 & a v_2 & a v_3 \end{vmatrix} = a \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = a\,(\mathsf{u} \times \mathsf{v});$$

$$\begin{aligned}
\mathsf{u}\times(\mathsf{v}+\mathsf{w}) &= \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ (v_1+w_1) & (v_2+w_2) & (v_3+w_3) \end{vmatrix} \\
&= \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} + \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & v_3 \end{vmatrix} \\
&= \mathsf{u}\times\mathsf{v}+\mathsf{u}\times\mathsf{w}.
\end{aligned}$$

These two properties show that $\mathbb{R}^3$ with the cross product form an algebra. This algebra is not commutative, since

$$\mathsf{u}\times\mathsf{v} = \begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = -\begin{vmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = -(\mathsf{v}\times\mathsf{u}).$$

In particular, notice that the equation above implies that $\mathsf{u}\times\mathsf{u}=\mathsf{0}$. Finally, this algebra is not associative as the following argument shows. First, from the definition of cross product it is simple to see that

$$\mathsf{e}_1\times\mathsf{e}_2=\mathsf{e}_3, \qquad \mathsf{e}_3\times\mathsf{e}_1=\mathsf{e}_2, \qquad \mathsf{e}_2\times\mathsf{e}_3=\mathsf{e}_1.$$

Using the first two equations we obtain the following relations,

$$\begin{aligned}
\mathsf{e}_1\times(\mathsf{e}_1\times\mathsf{e}_2) &= \mathsf{e}_1\times\mathsf{e}_3 = -\mathsf{e}_2, \\
(\mathsf{e}_1\times\mathsf{e}_1)\times\mathsf{e}_2 &= \mathsf{0}\times\mathsf{e}_2 = \mathsf{0},
\end{aligned}$$

showing that the algebra is not associative. ◁

In these notes we concentrate on only one example, the algebra of linear operators on a vector space, which we now describe in the following result.

Theorem 5.3.2. *The vector space $L(V)$ of linear operators on a vector space V over the field $\mathbb{F}$ equipped with the composition operation is an associative algebra with identity. This algebra is not commutative. Furthermore, $\dim L(V) = (\dim V)^2$.*

Proof of Theorem 5.3.2: The vector space $L(V)$ equipped with the composition operation is an algebra. Indeed, given arbitrary linear operators $\boldsymbol{R}$, $\boldsymbol{S}$, $\boldsymbol{T}\in L(V)$, the following equations holds for all $\boldsymbol{v}\in V$, and all scalars all a, $b\in\mathbb{F}$,

$$\begin{aligned}
\big(\boldsymbol{T}\circ(a\boldsymbol{S}+b\boldsymbol{R})\big)(\boldsymbol{v}) &= \boldsymbol{T}\big(a\,\boldsymbol{S}(\boldsymbol{v})+b\,\boldsymbol{R}(\boldsymbol{v})\big) \\
&= a\,\boldsymbol{T}\big(\boldsymbol{S}(\boldsymbol{v})\big)+b\,\boldsymbol{T}\big(\boldsymbol{R}(\boldsymbol{v})\big) \\
&= a\,\boldsymbol{T}\circ\boldsymbol{S}(\boldsymbol{v})+b\,\boldsymbol{T}\circ\boldsymbol{R}(\boldsymbol{v}).
\end{aligned}$$

Therefore, we established that the composition operation satisfies

$$\boldsymbol{T}\circ(a\boldsymbol{S}+b\boldsymbol{R}) = a\,\boldsymbol{T}\circ\boldsymbol{S}+b\,\boldsymbol{T}\circ\boldsymbol{R},$$

proving that the vector space $L(V)$ equipped with the composition operation is and algebra. This algebra is associative, since for all linear operators $\boldsymbol{R}$, $\boldsymbol{S}$, $\boldsymbol{T}\in L(V)$, the following equations holds for all $\boldsymbol{v}\in V$,

$$\begin{aligned}
\big(\boldsymbol{T}\circ(\boldsymbol{S}\circ\boldsymbol{R})\big)(\boldsymbol{v}) &= \boldsymbol{T}\big(\boldsymbol{S}\circ\boldsymbol{R}\,)(\boldsymbol{v})\big) \\
&= \boldsymbol{T}\big(\boldsymbol{S}\big(\boldsymbol{R}(\boldsymbol{v})\big)\big) \\
&= \big(\boldsymbol{T}\circ\boldsymbol{S}\big)\big(\boldsymbol{R}(\boldsymbol{v})\big) \\
&= \big(\big(\boldsymbol{T}\circ\boldsymbol{S}\big)\circ\boldsymbol{R}\big)(\boldsymbol{v}),
\end{aligned}$$

so we conclude that

$$\boldsymbol{T} \circ (\boldsymbol{S} \circ \boldsymbol{R}) = (\boldsymbol{T} \circ \boldsymbol{S}) \circ \boldsymbol{R}.$$

This algebra has identity, since the identity transformation $\boldsymbol{I}_V \in L(V)$ satisfies the equation $\boldsymbol{I}_V \circ \boldsymbol{T} = \boldsymbol{T} = \boldsymbol{T} \circ \boldsymbol{I}_V$ for all $\boldsymbol{T} \in L(V)$. Finally, this algebra is not commutative. We just need to show one example with this property. We choose $V = \mathbb{R}^2$ and the transformations given by matrices

$$\mathsf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

We have seen in Example 2.1.3 and 2.1.4 that A is a reflection along the axis $x_1 = x_2$ while B is a rotation by $\pi/2$ counterclockwise. It is simple to see that

$$\mathsf{A} \circ \mathsf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\mathsf{B} \circ \mathsf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Therefore, the algebra is not commutative. Finally, the result that $\dim L(V) = n^2$ follows from Theorem 5.2.9. This establishes the Theorem. □

EXAMPLE 5.3.2: Show that the vector space $\mathbb{F}^{n,n}$ equipped with the matrix multiplication operation is an associative algebra with identity.

SOLUTION: We know that $\mathbb{F}^{n,n} = L(\mathbb{F}^n)$, and we also know that matrix multiplication is equal to matrix composition, that is, for all $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{n,n}$ and all $\mathsf{x} \in \mathbb{F}^n$ holds

$$(\mathsf{AB})\mathsf{x} = \mathsf{A}(\mathsf{Bx}) = (\mathsf{A} \circ \mathsf{B})(\mathsf{x}) \quad \Rightarrow \quad \mathsf{AB} = \mathsf{A} \circ \mathsf{B}.$$

Therefore, Theorem 5.3.2 implies that $\mathbb{F}^{n,n}$ equipped with the matrix multiplication operation is an associative algebra with identity. ◁

5.3.1. **Polynomial functions of linear operators.** Polynomial functions of vectors in an algebra can be defined using linear combinations and multiplications of vectors. In the particular case of the algebra of linear operators, we use the notation $\boldsymbol{T} \circ \boldsymbol{S} = \boldsymbol{TS}$ for the composition of linear operators in $L(V)$, and also $\boldsymbol{T}^n = \boldsymbol{T}\boldsymbol{T}^{n-1}$ for the n-th power of the operator $\boldsymbol{T}$, which is defined for all positive integers $n \geqslant 1$. The consistency of this equation for $n = 1$ demands the definition $\boldsymbol{T}^0 = \boldsymbol{I}_V$. It follows that functions like the operator-valued polynomial below can be defined.

Definition 5.3.3. *An **operator-valued polynomial** of degree n defined by the scalars $a_0, \cdots, a_n \in \mathbb{F}$ is the function $p : L(V) \to L(V)$ given by*

$$p(\boldsymbol{T}) = a_0 \boldsymbol{I}_V + a_1 \boldsymbol{T} + a_2 \boldsymbol{T}^2 + \cdots + a_n \boldsymbol{T}^n.$$

EXAMPLE 5.3.3: Consider the vector space $V = \mathbb{R}^2$ and given any real number θ introduce the linear operator

$$\mathsf{R}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Find the explicit expression of the operator $(\mathsf{R}_\theta)^n$ for any positive integer n.

SOLUTION: We have seen in Example 2.1.5 that the operator R_θ is a rotation by an angle θ counterclockwise. We have also seen in Example 2.3.6 that given two operators R_{θ_1} and R_{θ_2}, the following formula holds,

$$\mathsf{R}_{\theta_1} \mathsf{R}_{\theta_2} = \mathsf{R}_{\theta_1 + \theta_2}.$$

In the particular case of $\theta_1 = \theta_2$ we obtain $\big(\mathsf{R}_\theta\big)^2 = \mathsf{R}_{2\theta}$. Analogously,

$$\big(\mathsf{R}_\theta\big)^3 = \mathsf{R}_\theta\big(\mathsf{R}_\theta\big)^2 = \mathsf{R}_\theta\mathsf{R}_{2\theta} = \mathsf{R}_{3\theta}.$$

This calculation suggests the formula $\big(\mathsf{R}_\theta\big)^n = \mathsf{R}_{n\theta}$ for all positive integer n. We now prove this formula using induction in n. Assume that the formula holds for an arbitrary value of n and then prove that the formula also holds for $(n+1)$. Indeed,

$$\big(\mathsf{R}_\theta\big)^{(n+1)} = \mathsf{R}_\theta\big(\mathsf{R}_\theta\big)^n = \mathsf{R}_\theta\mathsf{R}_{n\theta} = \mathsf{R}_{(n+1)\theta}.$$

We then conclude that $\big(\mathsf{R}_\theta\big)^n = \mathsf{R}_{n\theta}$ holds for every positive integer n. ◁

5.3.2. **Functions of linear operators.** Negative powers of an operator can be defined in the case that the operator is invertible. In that case, if one defines $\boldsymbol{T}^{-n} = \big(\boldsymbol{T}^{-1}\big)^n$ for any positive integer n, then the exponents satisfy the usual well known rules that hold when the base is a real number. That is, for an invertible operator $\boldsymbol{T}$ holds that $\boldsymbol{T}^m\boldsymbol{T}^n = \boldsymbol{T}^{(m+n)}$ and $\big(\boldsymbol{T}^m\big)^n = \boldsymbol{T}^{mn}$, for all integers m, n.

It is much more complicated to define further generalizations of these operator-valued power functions to include fractional exponents and real number exponents. The general idea behind such generalizations is the same used to define an infinitely differentiable function of an operator. Such functions are defined using Taylor expansions similar to those used on real-valued functions. Let us recall that the Taylor expansion centered at $x_0 \in \mathbb{R}$ of an infinitely differentiable real-valued function is given by

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}\,(x-x_0)^n = f(x_0) + f'(x_0)\,(x-x_0) + f''(x_0)\,(x-x_0)^2 + \cdots.$$

EXAMPLE 5.3.4: Find the Taylor series expansion of the real valued exponential function $f(x) = e^x$ centered $x_0 = 0$.

SOLUTION: Since $x_0 = 0$, the Taylor expansion formula above has the form

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}\,x^n = f(0) + f'(0)\,x + f''(0)\,x^2 + \cdots.$$

Since the exponential function $f(x) = e^x$ satisfies that $f^{(n)}(x) = f(x)$ for all positive integer n, and $f(0) = e^0 = 1$, we obtain the formula

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.$$

◁

Coming back to linear operators on a vector space, one can use the right hand side of the Taylor expansion formulas to define a function of a linear operator. However, such idea cannot be applied to linear operators until one understands the meaning of an infinite sum of linear operators. We discuss these ideas in Chapter 9. What we can do right now is to truncate the Taylor series and define a particular type of polynomial functions of linear operators as follows. Given the real-valued function $f : \mathbb{R} \to \mathbb{R}$, a vector space V, and a linear operator $\boldsymbol{T} \in L(V)$, introduce the Taylor polynomial $f_N(\boldsymbol{T})$ for any positive integer N as follows

$$f_N(\boldsymbol{T}) = \sum_{n=0}^{N} \frac{f^{(n)}(x_0)}{n!}\,\big(\boldsymbol{T} - x_0\,\boldsymbol{I}_V\big)^n,$$

that is,

$$f_N(\boldsymbol{T}) = f(x_0)\boldsymbol{I}_V + f'(x_0)\,\big(\boldsymbol{T} - x_0\,\boldsymbol{I}_V\big) + \cdots + \frac{f^{(N)}(x_0)}{N!}\,\big(\boldsymbol{T} - x_0\,\boldsymbol{I}_V\big)^N.$$

The Taylor polynomial of degree N simplify in the case that $x_0 = 0$, that is

$$f_N(\boldsymbol{T}) = \sum_{n=0}^{N} \frac{f^{(n)}(0)}{n!}\, \boldsymbol{T}^{\,n} = f(0)\boldsymbol{I}_V + f'(0)\, \boldsymbol{T} + \cdots + \frac{f^{(N)}(0)}{N!}\, \boldsymbol{T}^{\,N}.$$

In order to define the limit of $f_N(\boldsymbol{T})$ as $N \to \infty$ it is needed a notion of distance between linear operators. We come back to this point in Chapter 9.

5.3.3. **The commutator of linear operators.** We know that the composition of linear maps depends on the order the operators appear. We say that function composition is not commutative. The commutator of two linear operators measures the lack of commutativity of these two operators.

Definition 5.3.4. *The* ***commutator****,* $[\boldsymbol{T}, \boldsymbol{S}]$*, of the operators* $\boldsymbol{T}$*,* $\boldsymbol{S} \in L(V)$ *is given by,*

$$[\boldsymbol{T}, \boldsymbol{S}] = \boldsymbol{TS} - \boldsymbol{ST}.$$

In the case where $L(V) = \mathbb{F}^{n,n}$ the commutator of linear operators defined above agrees with the definition of matrix commutators introduced at the end of Section 2.3. An immediate consequence of the definition above are the following properties.

Theorem 5.3.5. *For* $\boldsymbol{S}$*,* $\boldsymbol{T}$*,* $\boldsymbol{U} \in L(V)$ *and* a*,* $b \in \mathbb{F}$*, we have:*

(a) $[\boldsymbol{T}, \boldsymbol{S}] = -[\boldsymbol{S}, \boldsymbol{T}]$, *antisymmetry;*
(b) $[a\boldsymbol{T}, b\boldsymbol{S}] = ab\,[\boldsymbol{T}, \boldsymbol{S}]$, *linearity;*
(c) $[\boldsymbol{U}, (\boldsymbol{T} + \boldsymbol{S})] = [\boldsymbol{U}, \boldsymbol{T}] + [\boldsymbol{U}, \boldsymbol{S}]$, *linearity in the right entry;*
(d) $[(\boldsymbol{U} + \boldsymbol{T}), \boldsymbol{S}] = [\boldsymbol{U}, \boldsymbol{S}] + [\boldsymbol{T}, \boldsymbol{S}]$, *linearity in the left entry;*
(e) $[\boldsymbol{UT}, \boldsymbol{S}] = \boldsymbol{U}[\boldsymbol{T}, \boldsymbol{S}] + [\boldsymbol{U}, \boldsymbol{S}]\boldsymbol{T}$, *left derivation property;*
(f) $[\boldsymbol{U}, \boldsymbol{TS}] = \boldsymbol{T}[\boldsymbol{U}, \boldsymbol{S}] + [\boldsymbol{U}, \boldsymbol{T}]\boldsymbol{S}$, *right derivation property;*
(g) $\big[[\boldsymbol{S}, \boldsymbol{T}], \boldsymbol{U}\big] + \big[[\boldsymbol{U}, \boldsymbol{S}], \boldsymbol{T}\big] + \big[[\boldsymbol{T}, \boldsymbol{U}], \boldsymbol{S}\big] = 0$, *Jacobi identity.*

Proof of Theorem 5.3.5: Properties (a)-(d) and (g) follow from the definition of commutator in a straightforward way. We just show here the left derivation property:

$$\begin{aligned}
[\boldsymbol{UT}, \boldsymbol{S}] &= \boldsymbol{UTS} - \boldsymbol{SUT} \\
&= \boldsymbol{UTS} - \boldsymbol{SUT} + \boldsymbol{UST} - \boldsymbol{UST} \\
&= (\boldsymbol{UTS} - \boldsymbol{UST}) + (\boldsymbol{UST} - \boldsymbol{SUT}) \\
&= \boldsymbol{U}[\boldsymbol{T}, \boldsymbol{S}] + [\boldsymbol{U}, \boldsymbol{S}]\boldsymbol{T}.
\end{aligned}$$

The proof of the right derivation property is similar. This establishes the Theorem. □

The commutator of two linear operators in a vector space is an important concept in quantum mechanics, since it indicates how well two properties of the physical system described by these operators can be measured simultaneously. The Heisenberg uncertainty relations are statements about the commutators of linear operators.

5.3.4. **Exercises.**

5.3.1.- Let $\boldsymbol{T}$, $\boldsymbol{S} \in L(\mathbb{F}^2)$ be the linear operators given by

$$\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}, \ \boldsymbol{S}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} 0 \\ x_1 \end{bmatrix}.$$

Compute explicitly the operators

(a) $3\boldsymbol{T} - 2\boldsymbol{S}$;
(b) $\boldsymbol{T} \circ \boldsymbol{S}$ and $\boldsymbol{S} \circ \boldsymbol{T}$;
(c) $\boldsymbol{T}^2$ and $\boldsymbol{S}^2$.

5.3.2.- Find the dimension of the vector spaces $L(\mathbb{F}^4)$, $L(\mathbb{P}_2)$ and $L(\mathbb{F}^{3,2})$.

5.3.3.- Let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator

$$\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} 2x_1 \\ 4x_1 - x_2 \end{bmatrix}.$$

Find the operator $\boldsymbol{T}^{-2}$.

5.3.4.- Let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator

$$\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{bmatrix}.$$

Find the operator

$$p(\boldsymbol{T}) = \boldsymbol{T}^2 - 2\,\boldsymbol{T} - 3\,\boldsymbol{I}_2.$$

5.3.5.- Use the definition of the rotation matrix R_θ given in Example 5.3.2 and the formula $\big(\mathsf{R}_\theta\big)^n = \mathsf{R}_{n\theta}$ to explicitly verify that $\big(\mathsf{R}_\theta\big)^n \big(\mathsf{R}_\theta\big)^{-n} = \mathsf{I}_2$.

5.3.6.- Prove the properties (a), (b) and (c) in Theorem 5.3.5.

5.4. Transformation components

5.4.1. **The matrix of a linear transformation.** We have seen in Sections 2.1 and 5.1 that every $m \times n$ matrix defines a linear transformation between the vector spaces $\mathbb{F}^n$ and $\mathbb{F}^m$ equiped with standard bases. We now show that every linear transformation $T : V \to W$ between finite dimensional vector spaces V and W has associated a matrix T. The matrix associated with such linear transformation is not unique. In order to compute the matrix of a linear transformation a basis must be chosen both in V and in W. The matrix associated with a linear transformation depends on the choice of bases done in V and W. We can say that the matrix T of a linear transformation $T : V \to W$ are the components of T in the given bases for V and W, analogously to the components v of a vector $\boldsymbol{v} \in V$ in a basis for V studied in Section 4.4.

Definition 5.4.1. *Let both V and W be finite dimensional vector spaces over the field $\mathbb{F}$ with respective ordered bases $\mathcal{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ and $\mathcal{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)$, and let the map $[\,]_w : W \to \mathbb{F}^m$ be the coordinate map on W. The* ***matrix of the linear transformation $\boldsymbol{T} : V \to W$*** *in the ordered bases $\mathcal{V}$ and $\mathcal{W}$ is the $m \times n$ matrix*

$$\mathsf{T}_{vw} = \big[\, [\boldsymbol{T}(\boldsymbol{v}_1)]_w, \cdots, [\boldsymbol{T}(\boldsymbol{v}_n)]_w \,\big]. \tag{5.3}$$

Let us explain the notation in Eq. (5.3). Given any basis vector $\boldsymbol{v}_i \in \mathcal{V} \subset V$, where $i = 1, \cdots, n$, we know that $\boldsymbol{T}(\boldsymbol{v}_i)$ is a vector in W. Like any vector in a vector space, $\boldsymbol{T}(\boldsymbol{v}_i)$ can be decomposed in a unique way in terms of the basis vectors in $\mathcal{W}$, and following Sect. 4.4 we denote them as $[\boldsymbol{T}(\boldsymbol{v}_i)]_w$. When there is no possibility of confusion, we denote T_{vw} simply as T.

Example 5.4.1: Consider the vector spaces $V = W = \mathbb{R}^2$, both with the standard basis, that is, $\mathcal{S} = \left(\mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$. Let the linear operator $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$\left[\boldsymbol{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s\right)\right]_s = \begin{bmatrix} 3x_1 + 2x_2 \\ 4x_1 - x_2 \end{bmatrix}_s. \tag{5.4}$$

(a) Find the matrix T_{ss} associated with the linear operator above.

(b) Consider the ordered basis $\mathcal{U}$ for $\mathbb{R}^2$ given by $\mathcal{U} = \left(\mathsf{u}_{1s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathsf{u}_{2s} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)$. Find the matrix T_{uu} associated with the linear operator above.

Solution:

Part (a): The definition of T_{ss} implies $\mathsf{T}_{ss} = \big[[\boldsymbol{T}(\mathsf{e}_1)]_s, [\boldsymbol{T}(\mathsf{e}_2)]_s\big]$. From Eq. (5.4) we know that

$$\left[\boldsymbol{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_s\right)\right]_s = \begin{bmatrix} 3 \\ 4 \end{bmatrix}_s, \qquad \left[\boldsymbol{T}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}_s\right)\right]_s = \begin{bmatrix} 2 \\ -1 \end{bmatrix}_s.$$

Therefore,

$$\mathsf{T}_{ss} = \begin{bmatrix} 3 & 2 \\ 4 & -1 \end{bmatrix}.$$

Part (b): By definition $\mathsf{T}_{uu} = \big[\, [\boldsymbol{T}(\boldsymbol{u}_1)]_u, [\boldsymbol{T}(\boldsymbol{u}_2)]_u \,\big]$. Notice that from the definition of basis $\mathcal{U}$ we have $\mathsf{u}_{is} = [\boldsymbol{u}_i]_s$, the column vector form with the components of the basis vectors $\boldsymbol{u}_i$ in the standard basis $\mathcal{S}$. So, we can use the definition of $\boldsymbol{T}$ in Eq. (5.4), that is,

$$[\boldsymbol{T}(\boldsymbol{u}_1)]_s = \left[\boldsymbol{T}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}_s\right)\right]_s = \begin{bmatrix} 5 \\ 3 \end{bmatrix}_s, \qquad [\boldsymbol{T}(\boldsymbol{u}_2)]_s = \left[\boldsymbol{T}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}_s\right)\right]_s = \begin{bmatrix} -1 \\ -5 \end{bmatrix}_s.$$

The results above are vector components in the standard basis $\mathcal{S}$, so we need to translate these results into components on the $\mathcal{U}$ basis. We translate them in the usual way, solving a linear system for each of these two vectors:

$$\begin{bmatrix}5\\3\end{bmatrix}_s = y_1 \begin{bmatrix}1\\1\end{bmatrix}_s + y_2 \begin{bmatrix}-1\\1\end{bmatrix}_s, \qquad \begin{bmatrix}-1\\-5\end{bmatrix}_s = z_1 \begin{bmatrix}1\\1\end{bmatrix}_s + z_2 \begin{bmatrix}-1\\1\end{bmatrix}_s.$$

The solutions are the components we are looking for, since

$$[\boldsymbol{T}(\boldsymbol{u}_1)]_u = \begin{bmatrix}y_1\\y_2\end{bmatrix}_u, \qquad [\boldsymbol{T}(\boldsymbol{u}_2)]_u = \begin{bmatrix}z_1\\z_2\end{bmatrix}_u.$$

We can solve both systems above at the same time using the augmented matrix

$$\left[\begin{array}{cc|cc}1 & -1 & 5 & -1\\ 1 & 1 & 3 & -5\end{array}\right] \to \left[\begin{array}{cc|cc}1 & 0 & 4 & -3\\ 0 & 1 & -1 & -2\end{array}\right].$$

We have obtained that

$$[\boldsymbol{T}(\boldsymbol{u}_1)]_u = \begin{bmatrix}4\\-1\end{bmatrix}_u, \qquad [\boldsymbol{T}(\boldsymbol{u}_2)]_u = \begin{bmatrix}-3\\-2\end{bmatrix}_u.$$

So we conclude that

$$\mathsf{T}_{uu} = \begin{bmatrix}4 & -3\\ -1 & -2\end{bmatrix}.$$

◁

From the Definition 5.4.1 it is clear that the matrix associated with a linear transformation $\boldsymbol{T}: V \to W$ depends on the choice of bases $\mathcal{V}$ and $\mathcal{W}$ for the spaces V and W, respectively. In the case that $V = W$ the matrix of the linear operator $\boldsymbol{T}: V \to V$ usually means T_{vv}, that is, to choose the same basis $\mathcal{V}$ for the domain space V and the range space V. But this is not the only choice. One can choose different bases $\mathcal{V}$ and $\tilde{\mathcal{V}}$ for the domain and range spaces, respectively. In this case, the matrix associated with the linear operator $\boldsymbol{T}$ is $\mathsf{T}_{v\tilde{v}}$, that is,

$$\mathsf{T}_{v\tilde{v}} = \big[\,[\boldsymbol{T}(\boldsymbol{v}_1)]_{\tilde{v}}, \cdots, [\boldsymbol{T}(\boldsymbol{v}_n)]_{\tilde{v}}\,\big].$$

EXAMPLE 5.4.2: Consider the linear operator defined in Eq. (5.4) in Example 5.4.1. Find the associated matrices T_{us} and T_{su}, where $\mathcal{S}$ and $\mathcal{U}$ are the bases defined in that Example 5.4.1.

SOLUTION: The first matrix is simple to obtain, since

$$\mathsf{T}_{us} = \big[[\boldsymbol{T}(\boldsymbol{u}_1)]_s, [\boldsymbol{T}(\boldsymbol{u}_2)]_s\big]$$

and it is straightforward to compute

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix}1\\1\end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix}5\\3\end{bmatrix}_s, \qquad \Big[\boldsymbol{T}\Big(\begin{bmatrix}-1\\1\end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix}-1\\-5\end{bmatrix}_s,$$

so, we then conclude that

$$\mathsf{T}_{us} = \begin{bmatrix}5 & -1\\ 3 & -5\end{bmatrix}_{us}.$$

We now compute

$$\mathsf{T}_{su} = \big[[\boldsymbol{T}(\boldsymbol{e}_1)]_u, [\boldsymbol{T}(\boldsymbol{e}_2)]_u\big].$$

From the definition of $\boldsymbol{T}$ we know that

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix}1\\0\end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix}3\\4\end{bmatrix}_s, \qquad \Big[\boldsymbol{T}\Big(\begin{bmatrix}0\\1\end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix}2\\-1\end{bmatrix}_s.$$

The results are expressed in the standard basis $\mathcal{S}$, so we need to translate them into the $\mathcal{U}$ basis, as follows

$$\begin{bmatrix}3\\4\end{bmatrix}_s = y_1 \begin{bmatrix}1\\1\end{bmatrix}_s + y_2 \begin{bmatrix}-1\\1\end{bmatrix}_s, \qquad \begin{bmatrix}2\\-1\end{bmatrix}_s = z_1 \begin{bmatrix}1\\1\end{bmatrix}_s + z_2 \begin{bmatrix}-1\\1\end{bmatrix}_s.$$

The solutions are the components we are looking for, since

$$[\boldsymbol{T}(\boldsymbol{e}_1)]_u = \begin{bmatrix}y_1\\y_2\end{bmatrix}_u, \qquad [\boldsymbol{T}(\boldsymbol{e}_2)]_u = \begin{bmatrix}z_1\\z_2\end{bmatrix}_u.$$

We can solve both systems above at the same time using the augmented matrix

$$\left[\begin{array}{cc|cc}1 & -1 & 3 & 2\\1 & 1 & 4 & -1\end{array}\right] \to \left[\begin{array}{cc|cc}2 & 0 & 7 & 1\\0 & 2 & 1 & -3\end{array}\right].$$

We have obtained that

$$[\boldsymbol{T}(\boldsymbol{e}_1)]_u = \frac{1}{2}\begin{bmatrix}7\\1\end{bmatrix}_u, \qquad [\boldsymbol{T}(\boldsymbol{e}_2)]_u = \frac{1}{2}\begin{bmatrix}1\\-3\end{bmatrix}_u.$$

So we conclude that

$$\mathsf{T}_{su} = \frac{1}{2}\begin{bmatrix}7 & 1\\1 & -3\end{bmatrix}_{su}.$$

◁

5.4.2. **Action as matrix-vector product.** An important property of the matrix associated with a linear transformation $\boldsymbol{T}$ is that the action of the transformation onto a vector can be represented as the matrix-vector product between the transformation matrix T the vector components in the appropriate bases.

Theorem 5.4.2. *Let V and W be finite dimensional vector spaces with ordered bases $\mathcal{V}$ and $\mathcal{W}$, respectively. Let $\boldsymbol{T} : V \to W$ be a linear transformation with associated matrix T_{vw}. Then, the components of the vector $\boldsymbol{T}(\boldsymbol{x}) \in W$ in the basis $\mathcal{W}$ can be expressed as the matrix-vector product*

$$[\boldsymbol{T}(\boldsymbol{x})]_w = \mathsf{T}_{vw}\,\mathsf{x}_v,$$

where $\mathsf{x}_v = [\boldsymbol{x}]_v$ are the components of the vector $\boldsymbol{x} \in V$ in the basis $\mathcal{V}$.

Proof of Theorem 5.4.2: Given any vector $\boldsymbol{x} \in V$, the definition of its vector components in the basis $\mathcal{V} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ implies that

$$\mathsf{x}_v = \begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}_{\mathcal{V}} \quad \Leftrightarrow \quad \boldsymbol{x} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n.$$

Since $\boldsymbol{T}$ is a linear transformation we know that

$$\boldsymbol{T}(\boldsymbol{x}) = x_1\,\boldsymbol{T}(\boldsymbol{v}_1) + \cdots + x_n\,\boldsymbol{T}(\boldsymbol{v}_n).$$

The equation above holds in any basis of W, in particular in $\mathcal{W}$, that is,

$$\begin{aligned}[\boldsymbol{T}(\boldsymbol{x})]_w &= x_1\,[\boldsymbol{T}(\boldsymbol{v}_1)]_w + \cdots + x_n\,[\boldsymbol{T}(\boldsymbol{v}_n)]_w\\ &= \big[\,[\boldsymbol{T}(\boldsymbol{v}_1)]_w, \cdots [\boldsymbol{T}(\boldsymbol{v}_n)]_w\,\big]\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}_v\\ &= \mathsf{T}_{vw}\,\mathsf{x}_v.\end{aligned}$$

This establishes the Theorem. □

EXAMPLE 5.4.3: Consider the linear operator $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ given in Example 5.4.1 above together with the bases $\mathcal{S}$ and $\mathcal{U}$ defined in that example.

(a) Use the matrix vector product to express the action of the operator T when the standard basis $\mathcal{S}$ is used in both domain and range spaces $\mathbb{R}^2$.
(b) Use the matrix vector product to express the action of the operator $\boldsymbol{T}$ when the basis $\mathcal{U}$ is used in both domain and range spaces $\mathbb{R}^2$.

SOLUTION:

Part (a): We need to find $[\boldsymbol{T}(\boldsymbol{x})]_s$ and express the result using T_{ss}. We use the notation $\mathsf{x}_s = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s$, and we repeat the steps followed in the proof of Theorem 5.4.2.

$$[\boldsymbol{T}(\boldsymbol{x})]_s = [\boldsymbol{T}(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2)]_s = x_1\,[\boldsymbol{T}(\boldsymbol{e}_1)]_s + x_2\,[\boldsymbol{T}(\boldsymbol{e}_2)]_s = \big[[\boldsymbol{T}(\boldsymbol{e}_1)]_s, [\boldsymbol{T}(\boldsymbol{e}_2)]_s\big] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s,$$

so we conclude that $[\boldsymbol{T}(\boldsymbol{x})]_s = \mathsf{T}_{ss}\mathsf{x}_s$ and the matrix on the far right-hand side agrees with the matrix we found in the first half of Example 5.4.1, where we have found that

$$\mathsf{T}_{ss} = \begin{bmatrix} 3 & -2 \\ 4 & 1 \end{bmatrix}_{ss}.$$

Therefore, the action of $\boldsymbol{T}$ on $\boldsymbol{x}$ when expressed in the standard basis $\mathcal{S}$ is given by the following matrix-vector product,

$$[\boldsymbol{T}(\boldsymbol{x})]_s = \begin{bmatrix} 3 & 2 \\ 4 & -1 \end{bmatrix}_{ss} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s.$$

Part (b): We need to find $[\boldsymbol{T}(\boldsymbol{x})]_u$ and express the result using T_{uu}. We use the notation $\mathsf{x}_u = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}_u$, and we repeat the steps followed in the proof of Theorem 5.4.2.

$$[\boldsymbol{T}(\boldsymbol{x})]_u = [\boldsymbol{T}(\tilde{x}_1\boldsymbol{u}_1 + \tilde{x}_2\boldsymbol{u}_2)]_u = \tilde{x}_1\,[\boldsymbol{T}(\boldsymbol{u}_1)]_u + \tilde{x}_2\,[\boldsymbol{T}(\boldsymbol{u}_2)]_u = \big[[\boldsymbol{T}(\boldsymbol{u}_1)]_u, [\boldsymbol{T}(\boldsymbol{u}_2)]_u\big] \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}_u,$$

so we conclude that $[\boldsymbol{T}(\boldsymbol{x})]_u = \mathsf{T}_{uu}\mathsf{x}_u$. At the end of Example 5.4.1 we have found that

$$\mathsf{T}_{uu} = \begin{bmatrix} 4 & -3 \\ -1 & -2 \end{bmatrix}_{uu}.$$

Therefore, the action of $\boldsymbol{T}$ on $\boldsymbol{x}$ when expressed in the basis $\mathcal{U}$ is given by the following matrix-vector product,

$$[\boldsymbol{T}(\boldsymbol{x})]_u = \begin{bmatrix} 4 & -3 \\ -1 & -2 \end{bmatrix}_{uu} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}_u.$$

◁

EXAMPLE 5.4.4: Express the action of the differentiation transformation $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_2$, given by $\boldsymbol{D}(\boldsymbol{p})(x) = \dfrac{d\boldsymbol{p}}{dx}(x)$ as a matrix-vector product in the standard ordered bases

$$\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2,\ \boldsymbol{p}_3 = x^3\big) \subset \mathbb{P}_3,$$
$$\tilde{\mathcal{S}} = \big(\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = x,\ \boldsymbol{q}_2 = x^2\big) \subset \mathbb{P}_2.$$

SOLUTION: Following the Theorem 5.4.2 we only need to find the matrix $\mathsf{D}_{s\tilde{s}}$ associated with the transformation $\boldsymbol{D}$,

$$\mathsf{D}_{s\tilde{s}} = \big[[\boldsymbol{D}(\boldsymbol{p}_0)]_{\tilde{s}}, [\boldsymbol{D}(\boldsymbol{p}_1)]_{\tilde{s}}, [\boldsymbol{D}(\boldsymbol{p}_2)]_{\tilde{s}}, [\boldsymbol{D}(\boldsymbol{p}_3)]_{\tilde{s}}\big] \quad \Rightarrow \quad \mathsf{D}_{s\tilde{s}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}_{s\tilde{s}}.$$

Therefore, given any vector

$$\boldsymbol{p}(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \quad \Leftrightarrow \quad \mathsf{p}_s = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}_s$$

we obtain that $\boldsymbol{D}(\boldsymbol{p})(x) = a_1 + 2a_2 x + 3a_3 x^2$ is equivalent to

$$[\boldsymbol{D}(\boldsymbol{p})]_{\tilde{s}} = \mathsf{D}_{s\tilde{s}}\mathsf{p}_s \quad \Rightarrow \quad [\boldsymbol{D}(\boldsymbol{p})]_{\tilde{s}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}_{s\tilde{s}} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}_s \quad \Rightarrow \quad [\boldsymbol{D}(\boldsymbol{p})]_{\tilde{s}} = \begin{bmatrix} a_1 \\ 2a_2 \\ 3a_3 \end{bmatrix}_{\tilde{s}}.$$

◁

EXAMPLE 5.4.5: Express the action of the integration transformation $\boldsymbol{S} : \mathbb{P}_2 \to \mathbb{P}_3$, given by $\boldsymbol{S}(\boldsymbol{q})(x) = \int_0^x \boldsymbol{q}(t)\,dt$ as a matrix-vector product in the standard ordered bases

$$\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2,\ \boldsymbol{p}_3 = x^3\big) \subset \mathbb{P}_3,$$
$$\tilde{\mathcal{S}} = \big(\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = x,\ \boldsymbol{q}_2 = x^2\big) \subset \mathbb{P}_2.$$

SOLUTION: Following the Theorem 5.4.2 we only need to find the matrix $\mathsf{S}_{\tilde{s}s}$ associated with the transformation $\boldsymbol{S}$,

$$\mathsf{S}_{\tilde{s}s} = \big[[\boldsymbol{S}(\boldsymbol{q}_0)]_s, [\boldsymbol{S}(\boldsymbol{q}_1)]_s, [\boldsymbol{S}(\boldsymbol{q}_2)]_s\big] \quad \Rightarrow \quad \mathsf{S}_{\tilde{s}s} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}_{\tilde{s}s}.$$

Therefore, given any vector

$$\boldsymbol{q}(x) = a_0 + a_1 x + a_2 x^2 \quad \Leftrightarrow \quad \mathsf{q}_{\tilde{s}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}_{\tilde{s}}$$

we obtain that $\boldsymbol{S}(\boldsymbol{q})(x) = a_0\, x + \dfrac{a_1}{2}\, x^2 + \dfrac{a_2}{3}\, x^3$ is equivalent to

$$[\boldsymbol{S}(\boldsymbol{q})]_s = \mathsf{S}_{\tilde{s}s}\mathsf{q}_{\tilde{s}} \quad \Rightarrow \quad [\boldsymbol{S}(\boldsymbol{q})]_s = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}_{\tilde{s}s} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}_{\tilde{s}} \quad \Rightarrow \quad [\boldsymbol{S}(\boldsymbol{q})]_s = \begin{bmatrix} 0 \\ a_0 \\ a_1/2 \\ a_2/3 \end{bmatrix}_s.$$

◁

5.4.3. **Composition and matrix product.** The following formula relates the composition of linear transformations with the matrix product of their associated matrices.

Theorem 5.4.3. *Let U, V and W be finite-dimensional vector spaces with bases $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$, respectively. Let $\boldsymbol{T} : U \to V$ and $\boldsymbol{S} : V \to W$ be linear transformations with associated matrices T_{uv} and S_{vw}, respectively. Then, the composition $\boldsymbol{S} \circ \boldsymbol{T} : U \to W$ given by $(\boldsymbol{S} \circ \boldsymbol{T})(\boldsymbol{u}) = \boldsymbol{S}\big(\big(\boldsymbol{T}(\boldsymbol{u})\big)$, for all $\boldsymbol{u} \in U$, is a linear transformation and the associated matrix $(\mathsf{S} \circ \mathsf{T})_{uw}$ is given by the matrix product by*

$$(\mathsf{S} \circ \mathsf{T})_{uw} = \mathsf{S}_{vw}\mathsf{T}_{uv}.$$

Proof of Theorem 5.4.3: First show that the composition of two linear transformations $\boldsymbol{S}$ and $\boldsymbol{T}$ is a linear transformation. Given any $\boldsymbol{u}_1$, $\boldsymbol{u}_2 \in U$ and arbitrary scalars a and b holds

$$\begin{aligned}(\boldsymbol{S} \circ \boldsymbol{T})(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) &= \boldsymbol{S}\big(\boldsymbol{T}(a\boldsymbol{u}_1 + b\boldsymbol{u}_2)\big) \\ &= \boldsymbol{S}\big(a\,\boldsymbol{T}(\boldsymbol{u}_1) + b\,\boldsymbol{T}(\boldsymbol{u}_2)\big) \\ &= a\,\boldsymbol{S}\big(\boldsymbol{T}(\boldsymbol{u}_1)\big) + b\,\boldsymbol{S}\big(\big(\boldsymbol{T}(\boldsymbol{u}_2)\big) \\ &= a\,(\boldsymbol{S} \circ \boldsymbol{T})(\boldsymbol{u}_1) + b\,(\boldsymbol{S} \circ \boldsymbol{T})(\boldsymbol{u}_2).\end{aligned}$$

We now compute the matrix of the composition transformation. Denote the ordered basis in U as follows, $\mathcal{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$. Then compute

$$(\mathsf{S} \circ \mathsf{T})_{uw} = \big[\,\big[\boldsymbol{S}\big(\boldsymbol{T}(\boldsymbol{u}_1)\big)\big]_w, \cdots, \big[\boldsymbol{S}\big(\boldsymbol{T}(\boldsymbol{u}_n)\big)\big]_w\,\big].$$

The column i, with $i = 1, \cdots, n$, in the matrix above has the form

$$\big[\boldsymbol{S}\big(\boldsymbol{T}(\boldsymbol{u}_i)\big)\big]_w = \mathsf{S}_{vw}[\boldsymbol{T}(\boldsymbol{u}_i)]_v = \mathsf{S}_{vw}\mathsf{T}_{uv}\mathsf{u}_{iu}.$$

Therefore, we conclude that

$$(\mathsf{S} \circ \mathsf{T})_{uw} = \mathsf{S}_{vw}\mathsf{T}_{uv}.$$

This establishes the Theorem. □

EXAMPLE 5.4.6: Let $\mathcal{S}$ be the standard ordered basis of $\mathbb{R}^3$, and consider the linear transformations $\boldsymbol{T} : \mathbb{R}^3 \to \mathbb{R}^3$ and $\boldsymbol{S} : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\left[\boldsymbol{T}\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}_s\right)\right]_s = \begin{bmatrix}2x_1 - x_2 + 3x_3\\-x_1 + 2x_2 - 4x_3\\x_2 + 3x_3\end{bmatrix}_s, \quad \left[\boldsymbol{S}\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}_s\right)\right]_s = \begin{bmatrix}-x_1\\2x_2\\3x_3\end{bmatrix}_s$$

(a) Find a matrices T_{ss} and S_{ss}. By the way, Is $\boldsymbol{T}$ injective? Is $\boldsymbol{T}$ surjective?
(b) Find the matrix of the composition $\boldsymbol{T} \circ \boldsymbol{S} : \mathbb{R}^3 \to \mathbb{R}^3$ in standard ordered basis $\mathcal{S}$.

SOLUTION: Since there is only one ordered basis in this problem, we represent T_{ss} and S_{ss} simply by T and S, respectively.

Part (a): A straightforward calculation from the definitions

$$\mathsf{T} = \big[[\boldsymbol{T}(\boldsymbol{e}_1)]_s, [\boldsymbol{T}(\boldsymbol{e}_2)]_s, [\boldsymbol{T}(\boldsymbol{e}_3)]_s\big], \quad \mathsf{S} = \big[[\boldsymbol{S}(\boldsymbol{e}_1)]_s, [\boldsymbol{S}(\boldsymbol{e}_2)]_s, [\boldsymbol{S}(\boldsymbol{e}_3)]_s\big],$$

gives us the matrices associated with $\boldsymbol{T}$ and $\boldsymbol{S}$ in the standard ordered basis,

$$\mathsf{T} = \begin{bmatrix}2 & -1 & 3\\-1 & 2 & -4\\0 & 1 & 3\end{bmatrix}, \qquad \mathsf{S} = \begin{bmatrix}-1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 3\end{bmatrix}.$$

The information in matrix T is useful to find out whether the linear transformation $\boldsymbol{T}$ is injective and/or surjective. First, find the reduced echelon form of T,

$$\begin{bmatrix} 2 & -1 & 3 \\ -1 & 2 & -4 \\ 0 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 4 \\ 2 & -1 & 3 \\ 0 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 4 \\ 0 & 3 & 5 \\ 0 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 10 \\ 0 & 1 & 3 \\ 0 & 0 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We conclude that $N(\mathsf{T}) = \{\boldsymbol{0}\}$, which implies that $\boldsymbol{T}$ is injective. The relation

$$\dim N(\mathsf{T}) + \dim \mathrm{Col}(\mathsf{T}) = 3$$

together with $\dim N(\mathsf{T}) = 0$ imply that $\dim \mathrm{Col}(\mathsf{T}) = 3$, hence $\boldsymbol{T}$ is surjective.

Part (b): The matrix of the composition $\boldsymbol{T} \circ \boldsymbol{S}$ in the standard ordered basis $\mathcal{S}$ is the product TS, that is,

$$(\mathsf{T} \circ \mathsf{S})_{ss} = \mathsf{TS} = \begin{bmatrix} 2 & -1 & 3 \\ -1 & 2 & -4 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \Rightarrow \quad (\mathsf{T} \circ \mathsf{S})_{ss} = \begin{bmatrix} -2 & -2 & 9 \\ 1 & 4 & -12 \\ 0 & 2 & 9 \end{bmatrix}.$$

◁

5.4.4. **Exercises.**

5.4.1.- Consider $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ the linear operator given by

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix} x_1 + x_2 \\ -2x_1 + 4x_2 \end{bmatrix}_s,$$

where $\mathcal{S}$ denote the standard ordered basis of $\mathbb{R}^2$. Find the matrix T_{uu} associated with the linear operator $\boldsymbol{T}$ in and the ordered basis $\mathcal{U}$ of $\mathbb{R}^2$ given by

$$\Big([\boldsymbol{u}_1]_s = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_s, [\boldsymbol{u}_2]_s = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_s\Big).$$

5.4.2.- Let $\boldsymbol{T} : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_1 - x_3 \end{bmatrix}_s,$$

$\mathcal{S}$ is the standard ordered basis of $\mathbb{R}^3$.

(a) Find the matrix T_{uu} of the linear operator $\boldsymbol{T}$ in the ordered basis

$$\mathcal{U} = \Big(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_s, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_s, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_s\Big).$$

(b) Verify that $[\boldsymbol{T}(\boldsymbol{v})]_u = \mathsf{T}_{uu}\mathsf{v}_u$, where

$$\mathsf{v}_s = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}_s.$$

5.4.3.- Fix an $m \times n$ matrix A and define the linear transformation $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^m$ as follows: $[\boldsymbol{T}(\boldsymbol{x})]_{\tilde{s}} = \mathsf{A}\mathsf{x}_s$, where $\mathcal{S} \subset \mathbb{R}^n$ and $\tilde{\mathcal{S}} \subset \mathbb{R}^m$ are standard ordered bases. Show that $\mathsf{T}_{s\tilde{s}} = \mathsf{A}$.

5.4.4.- Find the matrix associated with the linear transformation $\boldsymbol{T} : \mathbb{P}_3 \to \mathbb{P}_2$,

$$\boldsymbol{T}(\boldsymbol{p})(x) = \frac{d^2\boldsymbol{p}}{dx^2}(x) - \frac{d\boldsymbol{p}}{dx}(x),$$

in the standard ordered bases of $\mathbb{P}_3$, $\mathbb{P}_2$.

5.4.5.- Find the matrices in the standard bases of $\mathbb{P}_3$ and $\mathbb{P}_2$ of the transformations

$$\boldsymbol{S} \circ \boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_3, \quad \boldsymbol{D} \circ \boldsymbol{S} : \mathbb{P}_2 \to \mathbb{P}_2,$$

that is, the composition of the differentiation and integration transformations, as defined in Sect. 5.1.

5.5. Change of basis

In this Section we summarize the main calculations needed to find the change of vector and linear transformation components under a change of basis. We provide simple formulas to compute this change efficiently.

5.5.1. **Vector components.** We have seen in Sect. 4.4 that every vector $\boldsymbol{v}$ in a finite dimensional vector space V with an ordered basis $\mathcal{V}$, can be expressed in a unique way as a linear combination of the basis elements, with $\mathsf{v}_v = [\boldsymbol{v}]_v$ denoting the coefficients in that linear combination. These components depend on the basis chosen in V. Given two different ordered bases $\mathcal{V}$, $\tilde{\mathcal{V}} \subset V$, the components v_v and $\mathsf{v}_{\tilde{v}}$ associated with a vector $\boldsymbol{v} \in V$ are in general different. We now use the matrix-vector product to find a simple formula relating v_v to $\mathsf{v}_{\tilde{v}}$.

Let us recall the following notation. Given an n-dimensional vector space V, let $\tilde{\mathcal{V}}$ and $\mathcal{V}$ be two ordered bases of V given by

$$\tilde{\mathcal{V}} = \left(\tilde{\boldsymbol{v}}_1, \cdots, \tilde{\boldsymbol{v}}_n\right) \quad \text{and} \quad \mathcal{V} = \left(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\right).$$

Let $\boldsymbol{I} : V \to V$ be the identity transformation, that is, $\boldsymbol{I}(\boldsymbol{v}) = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$, and introduce the *change of basis matrices*

$$\mathsf{I}_{\tilde{v}v} = \left[\tilde{\mathsf{v}}_{1v}, \cdots, \tilde{\mathsf{v}}_{nv}\right] \quad \text{and} \quad \mathsf{I}_{v\tilde{v}} = \left[\mathsf{v}_{1\tilde{v}}, \cdots, \mathsf{v}_{n\tilde{v}}\right],$$

where we denoted, as usual, $\tilde{\mathsf{v}}_{iv} = [\tilde{\boldsymbol{v}}_i]_v$ and $\mathsf{v}_{i\tilde{v}} = [\boldsymbol{v}_i]_{\tilde{v}}$, for $i = 1, \cdots, n$. Since the sets $\mathcal{V}$ and $\tilde{\mathcal{V}}$ are bases of V, the matrices $\mathsf{I}_{v\tilde{v}}$ and $\mathsf{I}_{\tilde{v}v}$ are invertible, and it is not difficult to show that $(\mathsf{I}_{v\tilde{v}})^{-1} = \mathsf{I}_{\tilde{v}v}$. Finally, introduce the following notation for the change of basis matrices,

$$\mathsf{P} = \mathsf{I}_{\tilde{v}v} \quad \text{and} \quad \mathsf{P}^{-1} = \mathsf{I}_{v\tilde{v}}.$$

Theorem 5.5.1. *Let V be a finite dimensional vector space, let $\tilde{\mathcal{V}}$ and $\mathcal{V}$ be two ordered bases of V, and let $P = \mathsf{I}_{\tilde{v}v}$ be the change of basis matrix. Then, the components $\mathsf{x}_{\tilde{v}}$ and x_v of any vector $\boldsymbol{x} \in V$ in the ordered bases $\tilde{\mathcal{V}}$ and $\mathcal{V}$, respectively, are related by the linear equation*

$$\mathsf{x}_{\tilde{v}} = \mathsf{P}^{-1}\mathsf{x}_v. \tag{5.5}$$

Remark: Eq. (5.5) is equivalent to the inverse equation $\mathsf{x}_v = \mathsf{P}\,\mathsf{x}_{\tilde{v}}$.

Proof of Theorem 5.5.1: Let V be an n-dimensional vector space with two ordered bases $\tilde{\mathcal{V}} = \left(\tilde{\boldsymbol{v}}_1, \cdots, \tilde{\boldsymbol{v}}_n\right)$ and $\mathcal{V} = \left(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\right)$. Given any vector $\boldsymbol{x} \in V$, then the definition of vector components $\mathsf{x}_v = [\boldsymbol{x}]_v$ in the basis $\mathcal{V}$ implies

$$\mathsf{x}_v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_v \quad \Leftrightarrow \quad \boldsymbol{x} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n.$$

Express the second equation above in terms of components in the ordered basis $\tilde{\mathcal{V}}$,

$$\mathsf{x}_{\tilde{v}} = x_1\,\mathsf{v}_{1\tilde{v}} + \cdots + x_n\,\mathsf{v}_{n\tilde{v}} = \left[\mathsf{v}_{1\tilde{v}}, \cdots, \mathsf{v}_{n\tilde{v}}\right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_v \quad \Rightarrow \quad \mathsf{x}_{\tilde{v}} = \mathsf{I}_{v\tilde{v}}\,\mathsf{x}_v.$$

We conclude that $\mathsf{x}_{\tilde{v}} = \mathsf{P}^{-1}\mathsf{x}_v$. This establishes the Theorem. □

EXAMPLE 5.5.1: Consider the vector space $V = \mathbb{R}^2$ with the standard ordered basis $\mathcal{S}$ and the ordered basis $\mathcal{U} = \left(\mathsf{u}_{1s} = \begin{bmatrix}1\\1\end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix}-1\\1\end{bmatrix}_s\right)$. Given the vector with components $\mathsf{x}_s = \begin{bmatrix}1\\3\end{bmatrix}_s$, find x_u.

SOLUTION: The answer is given by Theorem 5.5.1, that says

$$\mathsf{x}_u = \mathsf{P}^{-1}\mathsf{x}_s, \qquad \mathsf{P} = \mathsf{I}_{us}.$$

From the data of the problem the matrix P is simple to compute, since

$$\mathsf{P} = \mathsf{I}_{us} = \left[\mathsf{u}_{1s}, \mathsf{u}_{2s}\right] = \begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}_{us}.$$

Computing the inverse matrix

$$\mathsf{P}^{-1} = \frac{1}{2}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}_{su}$$

we obtain the final result

$$\mathsf{x}_u = \frac{1}{2}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}_{su}\begin{bmatrix}1\\3\end{bmatrix}_s \quad \Rightarrow \quad \mathsf{x}_u = \begin{bmatrix}2\\1\end{bmatrix}_u.$$

◁

EXAMPLE 5.5.2: Let $V = \mathbb{R}^2$ with ordered bases $\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ and $\mathcal{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2\}$ related by the equations

$$\boldsymbol{b}_1 = -\boldsymbol{c}_1 + 4\boldsymbol{c}_2, \qquad \boldsymbol{b}_2 = 5\boldsymbol{c}_1 - 3\boldsymbol{c}_2.$$

(a) Given $\mathsf{x}_b = \begin{bmatrix}5\\3\end{bmatrix}_b$, find x_c.

(b) Given $\mathsf{x}_c = \begin{bmatrix}1\\1\end{bmatrix}_c$, find x_b.

SOLUTION:

Part (a): We know that $\mathsf{x}_c = \mathsf{P}^{-1}\mathsf{x}_b$, where $\mathsf{P} = \mathsf{I}_{cb}$. From the data of the problem we know that

$$\left.\begin{aligned}\boldsymbol{b}_1 = -\boldsymbol{c}_1 + 4\boldsymbol{c}_2 \quad &\Leftrightarrow \quad \mathsf{b}_{1c} = \begin{bmatrix}-1\\4\end{bmatrix}_c,\\ \boldsymbol{b}_2 = 5\boldsymbol{c}_1 - 3\boldsymbol{c}_2. \quad &\Leftrightarrow \quad \mathsf{b}_{2c} = \begin{bmatrix}5\\-3\end{bmatrix}_c,\end{aligned}\right\} \quad \Rightarrow \quad \mathsf{I}_{bc} = \begin{bmatrix}-1 & 5\\ 4 & -3\end{bmatrix}_{bc},$$

hence we know the change of basis matrix $\mathsf{I}_{bc} = \mathsf{P}^{-1}$, so we conclude that

$$\mathsf{x}_c = \begin{bmatrix}-1 & 5\\ 4 & -3\end{bmatrix}_{bc}\begin{bmatrix}5\\3\end{bmatrix}_b \quad \Rightarrow \quad \mathsf{x}_c = \begin{bmatrix}10\\11\end{bmatrix}_c$$

Part (b): keeping the definition of matrix $\mathsf{P} = \mathsf{I}_{cb}$ as we introduced it in part (a), we know that $\mathsf{x}_c = \mathsf{P}^{-1}\mathsf{x}_b$. In this part (b) we need the inverse relation $\mathsf{x}_b = \mathsf{P}\mathsf{x}_c$. Since $\mathsf{P}^{-1} = \begin{bmatrix}-1 & 5\\ 4 & -3\end{bmatrix}_{bc}$, it is simple to obtain $\mathsf{P}^{-1} = \frac{1}{17}\begin{bmatrix}3 & 5\\ 4 & 1\end{bmatrix}_{cb}$. Using this matrix we obtain

$$\mathsf{x}_b = \frac{1}{17}\begin{bmatrix}3 & 5\\ 4 & 1\end{bmatrix}_{cb}\begin{bmatrix}1\\1\end{bmatrix}_c \quad \Rightarrow \quad \mathsf{x}_b = \frac{1}{17}\begin{bmatrix}8\\5\end{bmatrix}_b.$$

◁

5.5.2. **Transformation components.** Analogously, we have seen in Sect. 5.4 that every linear transformation $\boldsymbol{T}: V \to W$ between finite dimensional vector spaces V and W with ordered bases $\mathcal{V}$ and $\mathcal{W}$, respectively, can be expressed in a unique way as an $\dim W \times \dim V$ matrix T_{vw}. These components depend on the bases chosen in V and W. Given two different ordered bases $\mathcal{V}, \tilde{\mathcal{V}} \subset V$, and given two different bases $\mathcal{W}, \tilde{\mathcal{W}} \subset W$, the matrices T_{vw} and $\mathsf{T}_{\tilde{v}\tilde{w}}$ associated with the linear transformation $\boldsymbol{T}$ are in general different. Matrix multiplication provides a simple formula relating T_{vw} and $\mathsf{T}_{\tilde{v}\tilde{w}}$.

Let us recall old notation and also introduce a bit of new one. Given an n-dimensional vector space V, let $\tilde{\mathcal{V}}$ and $\mathcal{V}$ be ordered bases of V,

$$\tilde{\mathcal{V}} = \big(\tilde{\boldsymbol{v}}_1, \cdots, \tilde{\boldsymbol{v}}_n\big) \quad \text{and} \quad \mathcal{V} = \big(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\big);$$

and given an m-dimensional vector space W, let $\tilde{\mathcal{W}}$ and $\mathcal{W}$ be two ordered bases of W,

$$\tilde{\mathcal{W}} = \big(\tilde{\boldsymbol{w}}_1, \cdots, \tilde{\boldsymbol{w}}_m\big) \quad \text{and} \quad \mathcal{W} = \big(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\big).$$

Let $\boldsymbol{I}: V \to V$ be the identity transformation, that is, $\boldsymbol{I}(\boldsymbol{v}) = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$, and introduce the change of basis matrices

$$\mathsf{I}_{v\tilde{v}} = \big[\mathsf{v}_{1\tilde{v}}, \cdots, \mathsf{v}_{n\tilde{v}}\big] \quad \text{and} \quad \mathsf{I}_{\tilde{v}v} = \big[\tilde{\mathsf{v}}_{1v}, \cdots, \tilde{\mathsf{v}}_{nv}\big].$$

Let $\boldsymbol{J}: W \to W$ be the identity transformation, that is, $\boldsymbol{J}(\boldsymbol{w}) = \boldsymbol{w}$ for all $\boldsymbol{w} \in W$, and introduce the change of basis matrices

$$\mathsf{J}_{w\tilde{w}} = \big[\mathsf{w}_{1\tilde{w}}, \cdots, \mathsf{w}_{m\tilde{w}}\big] \quad \text{and} \quad \mathsf{J}_{\tilde{w}w} = \big[\tilde{\mathsf{w}}_{1w}, \cdots, \tilde{\mathsf{w}}_{mw}\big].$$

Notice that the sets $\mathcal{V}$ and $\tilde{\mathcal{V}}$ are bases of V, therefore the $n \times n$ matrices $\mathsf{I}_{v\tilde{v}}$ and $\mathsf{I}_{\tilde{v}}$ are invertible, and $(\mathsf{I}_{v\tilde{v}})^{-1} = \mathsf{I}_{\tilde{v}v}$. The similar statement is true for the $m \times m$ matrices $\mathsf{J}_{w\tilde{w}}$ and $\mathsf{J}_{\tilde{w}w}$, and $(\mathsf{J}_{w\tilde{w}})^{-1} = \mathsf{J}_{\tilde{w}w}$. Finally, introduce the following notation for the change of basis matrices,

$$\mathsf{P} = \mathsf{I}_{\tilde{v}v} \quad \Rightarrow \quad \mathsf{P}^{-1} = \mathsf{I}_{v\tilde{v}}, \quad \text{and} \quad \mathsf{Q} = \mathsf{J}_{\tilde{w}w} \quad \Rightarrow \quad \mathsf{Q}^{-1} = \mathsf{J}_{w\tilde{w}}.$$

Theorem 5.5.2. *Let V and W be finite dimensional vector spaces, let $\tilde{\mathcal{V}}$ and $\mathcal{V}$ be two ordered bases of V, let $\tilde{\mathcal{W}}$ and $\mathcal{W}$ be two ordered bases of W, and let $\mathsf{P} = \mathsf{I}_{\tilde{v}v}$ and $\mathsf{Q} = \mathsf{J}_{\tilde{w}w}$ be the change of basis matrices, respectively. Then, the components $\mathsf{T}_{\tilde{v}\tilde{w}}$ and T_{vw} of any linear transformation $\boldsymbol{T}: V \to W$ in the bases $\tilde{\mathcal{V}}$, $\tilde{\mathcal{W}}$ and $\mathcal{V}$, $\mathcal{W}$, respectively, are related by the matrix equation*

$$\mathsf{T}_{\tilde{v}\tilde{w}} = \mathsf{Q}^{-1}\mathsf{T}_{vw}\mathsf{P}. \tag{5.6}$$

REMARK: Eq. (5.6) is equivalent to the inverse equation $\mathsf{T}_{vw} = \mathsf{Q}\mathsf{T}_{\tilde{v}\tilde{w}}\mathsf{P}^{-1}$. A particular case of Theorem 5.5.2 that frequently appears in applications is when T is a linear operator. In this is the case $\boldsymbol{T}: V \to V$, so $V = W$. Often in applications one also has $\mathcal{V} = \mathcal{W}$ and $\tilde{\mathcal{V}} = \tilde{\mathcal{W}}$, which imply $\mathsf{P} = \mathsf{Q}$.

Corollary 5.5.3. *Let V be a finite dimensional vector space, let $\tilde{\mathcal{V}}$ and $\mathcal{V}$ be two ordered bases of V, let $\mathsf{P} = \mathsf{I}_{\tilde{v}v}$ be the change of basis matrix, and let $\boldsymbol{T}: V \to V$ be a linear operator. Then, the components $\tilde{\mathsf{T}} = \mathsf{T}_{\tilde{v}\tilde{v}}$ and $\mathsf{T} = \mathsf{T}_{vv}$ of $\boldsymbol{T}$ in the bases $\tilde{\mathcal{V}}$ and $\mathcal{V}$, respectively, are related by the matrix equation*

$$\tilde{\mathsf{T}} = \mathsf{P}^{-1}\mathsf{T}\mathsf{P}. \tag{5.7}$$

Proof of Theorem 5.5.2: Let V be an n-dimensional vector space with ordered bases $\tilde{\mathcal{V}} = \big(\tilde{\boldsymbol{v}}_1, \cdots, \tilde{\boldsymbol{v}}_n\big)$ and $\mathcal{V} = \big(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\big)$; and let W be an m-dimensional vector space with ordered bases $\tilde{\mathcal{W}} = \big(\tilde{\boldsymbol{w}}_1, \cdots, \tilde{\boldsymbol{w}}_m\big)$ and $\mathcal{W} = \big(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\big)$. We know that the matrix $\mathsf{T}_{\tilde{v}\tilde{w}}$ associated with the transformation $\boldsymbol{T}$ and the bases $\tilde{\mathcal{V}}$, $\tilde{\mathcal{W}}$, and the matrix T_{vw} associated with the transformation $\boldsymbol{T}$ and the bases $\mathcal{V}$, $\mathcal{W}$ satisfy the following equations,

$$[\boldsymbol{T}(\boldsymbol{x})]_{\tilde{w}} = \mathsf{T}_{\tilde{v}\tilde{w}}\mathsf{x}_{\tilde{v}} \quad \text{and} \quad [\boldsymbol{T}(\boldsymbol{x})]_w = \mathsf{T}_{vw}\mathsf{x}_v. \tag{5.8}$$

Since $\boldsymbol{T}(\boldsymbol{x}) \in W$, Theorem 5.5.1 says that its components in the bases $\tilde{\mathcal{W}}$ and $\mathcal{W}$ are related by the equation

$$[\boldsymbol{T}(\boldsymbol{x})]_{\tilde{w}} = \mathsf{Q}^{-1}\,[\boldsymbol{T}(\boldsymbol{x})]_w, \quad \text{with} \quad \mathsf{Q} = \mathsf{I}_{\tilde{w}w}.$$

Therefore, the first equation in (5.8) implies that

$$\begin{aligned} \mathsf{T}_{\tilde{v}\tilde{w}}\mathsf{x}_{\tilde{v}} &= \mathsf{Q}^{-1}[\boldsymbol{T}(\boldsymbol{x})]_w \\ &= \mathsf{Q}^{-1}\mathsf{T}_{vw}\mathsf{x}_v \\ &= \mathsf{Q}^{-1}\mathsf{T}_{vw}\mathsf{P}\,\mathsf{x}_{\tilde{v}}, \quad \text{with} \quad \mathsf{P} = \mathsf{I}_{\tilde{v}v}, \end{aligned}$$

where in the second line above we used the second equation in (5.8), and in the third line above we used the inverse form of Theorem 5.5.1. Since the equation above holds for all $\boldsymbol{x} \in V$ we conclude that

$$\mathsf{T}_{\tilde{v}\tilde{w}} = \mathsf{Q}^{-1}\mathsf{T}_{vw}\mathsf{P}.$$

This establishes the Theorem. □

EXAMPLE 5.5.3: Let $\mathcal{S}$ be the standard ordered basis of $\mathbb{R}^2$, and let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator given by

$$\left[\boldsymbol{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s\right)\right]_s = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}_s,$$

that is, a reflection along the line $x_2 = x_1$. Find the matrix T_{uu}, where the ordered basis $\mathcal{U}$ is given by $\mathcal{U} = \left(\mathsf{u}_{1s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}_s\right)$.

SOLUTION: From the definition of $\boldsymbol{T}$ is straightforward to obtain T_{ss}, as follows,

$$\mathsf{T}_{ss} = \left[\left[\boldsymbol{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_s\right)\right]_s, \left[\boldsymbol{T}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}_s\right)\right]_s\right] \quad \Rightarrow \quad \mathsf{T}_{ss} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{ss}.$$

Since $\boldsymbol{T}$ is a linear operator and we want to compute the matrix T_{uu} from matrix T_{ss}, we can use Corollary 5.5.3, which says that these matrices are related by the similarity transformation

$$\mathsf{T}_{uu} = \mathsf{P}^{-1}\mathsf{T}_{ss}\mathsf{P}, \quad \text{where} \quad \mathsf{P} = \mathsf{I}_{us}.$$

From the data of the problem we know that

$$\mathsf{I}_{us} = \left[\mathsf{u}_{1s}, \mathsf{u}_{2s}\right] us = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}_{us},$$

therefore,

$$\mathsf{P} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}_{us}, \qquad \mathsf{P}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}_{su}.$$

We then conclude that

$$\mathsf{T}_{uu} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}_{su}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{ss}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}_{us} \quad \Rightarrow \quad \mathsf{T}_{uu} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_{uu}.$$

REMARK: We can see in this Example that the matrix associated with the reflection transformation T is diagonal in the basis $\mathcal{U}$. For this particular transformation we have that

$$\boldsymbol{T}(\boldsymbol{u}_1) = \boldsymbol{u}_1, \qquad \boldsymbol{T}(\boldsymbol{u}_2) = -\boldsymbol{u}_2.$$

Non-zero vectors $\boldsymbol{v}$ with this property, that $\boldsymbol{T}(\boldsymbol{v}) = \lambda\boldsymbol{v}$, are called eigenvectors of the operator $\boldsymbol{T}$ and the scalar λ is called eigenvalue of $\boldsymbol{T}$. In this example the elements in the basis $\mathcal{U}$ are eigenvectors of the reflection operator, and the matrix of $\boldsymbol{T}$ in this special basis is diagonal. Basis formed with eigenvectors of a given linear operator will be studied in a later on. ◁

EXAMPLE 5.5.4: Let $\mathcal{S}$ be the standard ordered basis of $\mathbb{R}^2$, and let $\boldsymbol{T}: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator given by

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s\Big)\Big]_s = \frac{(x_1 + x_2)}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}_s,$$

that is, a projection along the line $x_2 = x_1$. Find the matrix T_{uu}, where the ordered basis $\mathcal{U} = \Big(\mathsf{u}_{1s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}_s\Big)$.

SOLUTION: From the definition of $\boldsymbol{T}$ is straightforward to obtain T_{ss}, as follows,

$$\mathsf{T}_{ss} = \left[\Big[\boldsymbol{T}\Big(\begin{bmatrix} 1 \\ 0 \end{bmatrix}_s\Big)\Big]_s, \Big[\boldsymbol{T}\Big(\begin{bmatrix} 0 \\ 1 \end{bmatrix}_s\Big)\Big]_s\right] \quad \Rightarrow \quad \mathsf{T}_{ss} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}_{ss}.$$

Since $\boldsymbol{T}$ is a linear operator and we want to compute the matrix T_{uu} from matrix T_{ss}, we again use Corollary 5.5.3, which says that these matrices are related by the similarity transformation

$$\mathsf{T}_{uu} = \mathsf{P}^{-1}\mathsf{T}_{ss}\mathsf{P}, \quad \text{where} \quad \mathsf{P} = \mathsf{I}_{us}.$$

In Example 5.5.3 we have already computed

$$\mathsf{P} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}_{us}, \qquad \mathsf{P}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}_{su}.$$

We then conclude that

$$\mathsf{T}_{uu} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}_{su} \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}_{ss} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}_{us} \quad \Rightarrow \quad \mathsf{T}_{uu} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}_{uu}.$$

REMARK: Again in this Example we can see that the matrix associated with the reflection transformation $\boldsymbol{T}$ is diagonal in the basis $\mathcal{U}$, with diagonal elements equal to one and zero. For this particular transformation we have that the basis vectors in $\mathcal{U}$ are eigenvectors of $\boldsymbol{T}$ with eigenvalues 1 and 0, that is, $\boldsymbol{T}(\boldsymbol{u}_1) = \boldsymbol{u}_1$ and $\boldsymbol{T}(\boldsymbol{u}_2) = \boldsymbol{0}$. ◁

EXAMPLE 5.5.5: Let $\mathcal{U}$ and $\mathcal{V}$ be standard ordered bases of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively, let $\boldsymbol{T}: \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_u\Big)\Big]_v = \begin{bmatrix} x_1 - x_2 + x_3 \\ x_2 - x_3 \end{bmatrix}_v,$$

and introduce the ordered bases

$$\tilde{\mathcal{U}} = \Big(\tilde{\mathsf{u}}_{1u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_u, \tilde{\mathsf{u}}_{2u} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_u, \tilde{\mathsf{u}}_{3u} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_u\Big) \subset \mathbb{R}^3,$$

$$\tilde{\mathcal{V}} = \Big(\tilde{\mathsf{v}}_{1v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_v, \tilde{\mathsf{v}}_{2v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}_v\Big) \subset \mathbb{R}^2.$$

Find the matrices T_{uv} and $\mathsf{T}_{\tilde{u}\tilde{v}}$.

SOLUTION: We start finding the matrix T_{uv}, which by definition is given by

$$\mathsf{T}_{uv} = \left[\Big[\boldsymbol{T}\Big(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_u\Big)\Big]_v, \Big[\boldsymbol{T}\Big(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_u\Big)\Big]_v, \Big[\boldsymbol{T}\Big(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_u\Big)\Big]_v\right]$$

hence we obtain $\mathsf{T}_{uv} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix}_{uv}$. Theorem 5.5.2 says that the matrices $\mathsf{T}_{\tilde{u}\tilde{v}}$ and T_{uv} are related by the equation

$$\mathsf{T}_{\tilde{u}\tilde{v}} = \mathsf{Q}^{-1}\mathsf{T}_{uv}\mathsf{P}, \quad \text{where} \quad \mathsf{Q} = \mathsf{J}_{\tilde{v}v} \quad \text{and} \quad \mathsf{P} = \mathsf{I}_{\tilde{u}u}.$$

From the data of the problem we know that

$$\mathsf{I}_{\tilde{u}u} = \left[\tilde{\mathsf{u}}_{1u}, \tilde{\mathsf{u}}_{2u}, \tilde{\mathsf{u}}_{3u}\right]_{\tilde{u}u} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}_{\tilde{u}u} \quad \Rightarrow \quad \mathsf{P} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}_{\tilde{u}u},$$

$$\mathsf{J}_{\tilde{v}v} = \left[\tilde{\mathsf{v}}_{1v}, \tilde{\mathsf{v}}_{2v}\right]_{\tilde{v}v} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}_{\tilde{v}v} \quad \Rightarrow \quad \mathsf{Q} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}_{\tilde{v}v}, \quad \mathsf{Q}^{-1} = \frac{1}{3}\begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}_{v\tilde{v}}.$$

Therefore, we need to compute the matrix product

$$\mathsf{T}_{\tilde{u}\tilde{v}} = \frac{1}{3}\begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}_{v\tilde{v}} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix}_{uv} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}_{\tilde{u}u}$$

and the result is $\mathsf{T}_{\tilde{u}\tilde{v}} = \frac{1}{3}\begin{bmatrix} 2 & 0 & -4 \\ -1 & 0 & 5 \end{bmatrix}_{\tilde{u}\tilde{v}}$. ◁

5.5.3. **Determinant and trace of linear operators.** The type of matrix transformation given by Eq. (5.7) will be important later on, so we give such transformations a name.

Definition 5.5.4. *The $n \times n$ matrices A and B are related by a* ***similarity transformation*** *iff there exists and invertible $n \times n$ matrix P such that*

$$\mathsf{B} = \mathsf{P}^{-1}\mathsf{A}\mathsf{P}.$$

Therefore, similarity transformations are transformations among matrices. They appear in two different contexts. The first, or *active* context, is when matrix A and matrix B above are written in the same basis. In this case a similarity transformation changes matrix A into matrix B. The second, or *passive* context, is when matrix A and matrix B are two matrices corresponding to the same linear transformation. In this case matrices A and B are written in different basis, and the similarity transformation is the change of basis equation for the linear transformation matrices.

The following results says that, no matter in which contex one uses a similarity transformation, the determinant and trace of a matrix are invariant under similarity transformations.

Theorem 5.5.5. *If the square matrices A, B are related by the similarity transformation $\mathsf{B} = \mathsf{P}^{-1}\mathsf{A}\mathsf{P}$, then holds*

$$\det(\mathsf{B}) = \det(\mathsf{A}), \qquad \text{tr}\,(\mathsf{B}) = \text{tr}\,(\mathsf{A}).$$

Proof of Theorem: 5.5.5: The determinant is invariant under similarity transformations, since

$$\det(\mathsf{B}) = \det(\mathsf{P}^{-1}\mathsf{A}\mathsf{P}) = \det(\mathsf{P}^{-1})\det(\mathsf{A})\det(\mathsf{P}) = \det(\mathsf{A}).$$

The trace is also invariant under similarity transformations, since

$$\text{tr}\,(\mathsf{B}) = \text{tr}\,(\mathsf{P}^{-1}\mathsf{A}\mathsf{P}) = \text{tr}\,(\mathsf{P}\mathsf{P}^{-1}\mathsf{A}) = \text{tr}\,(\mathsf{A}).$$

This establishes the Theorem. □

We now use this result in the passive context for similarity transformations. This means that the determinant and trace are independent of the vector basis one uses to write down the matrix. Therefore, a determinant or a trace is an operation on a linear transformation, not on a particular matrix representation of a linear transformation. This observation suggests the following definition.

Definition 5.5.6. *Let V be a finite dimensional vector space, let $\boldsymbol{T} \in L(V)$ be a linear operator, and let T_{vv} be the matrix of the linear operator in an arbitrary ordered basis $\mathcal{V}$ of V. The determinant and trace of a linear operator $\boldsymbol{T} \in L(V)$ are respectively given by,*

$$\det(\boldsymbol{T}) = \det(\mathsf{T}_{vv}), \qquad \mathrm{tr}\,(\boldsymbol{T}) = \mathrm{tr}\,(\mathsf{T}_{vv}).$$

We repeat that the determinant and trace of a linear operator are well-defined, since given any other ordered basis $\mathcal{V}$ of V, we know that the matrices $\mathsf{T}_{\tilde{v}\tilde{v}}$ and T_{vv} are related by a similarity transformation. However, the determinant and trace are invariant under similatiry transformations. So no matter what vector basis we use to compute determinant and trace of a linear operator, we always get the same result.

5.5.4. **Exercises.**

5.5.1.- Let $\mathcal{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$ be an ordered basis of $\mathbb{R}^2$ given by

$$\boldsymbol{u}_1 = 2\boldsymbol{e}_1 - 9\boldsymbol{e}_2, \quad \boldsymbol{u}_2 = \boldsymbol{e}_1 + 8\boldsymbol{e}_2,$$

where $\mathcal{S} = (\mathbf{e}_1, \mathbf{e}_2)$ is the standard ordered basis of $\mathbb{R}^2$.

(a) Find both change of basis matrices I_{us} and I_{su}.

(b) Given the vector $\boldsymbol{x} = 2\boldsymbol{u}_1 + \boldsymbol{u}_2$, find both x_s and x_u.

5.5.2.- Consider the ordered bases of $\mathbb{R}^3$, $\mathcal{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ and $\mathcal{C} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3)$, where

$$\boldsymbol{c}_1 = \boldsymbol{b}_1 - 2\boldsymbol{b}_2 + \boldsymbol{b}_3,$$
$$\boldsymbol{c}_2 = -\boldsymbol{b}_2 + 3\boldsymbol{b}_3,$$
$$\boldsymbol{c}_3 = -2\boldsymbol{b}_1 + \boldsymbol{b}_3.$$

(a) Find both the change of basis matrices I_{bc} and I_{cb}.

(b) Let $\boldsymbol{x} = \boldsymbol{c}_1 - 2\boldsymbol{c}_2 + 2\boldsymbol{c}_3$. Find both x_b and x_c.

5.5.3.- Consider the ordered bases $\mathcal{U}$ and $\mathcal{S}$ of $\mathbb{R}^2$ given by

$$\mathcal{U} = \Big(\mathsf{u}_{1s} = \begin{bmatrix}1\\2\end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix}2\\1\end{bmatrix}_s\Big),$$

$$\mathcal{S} = \Big(\mathsf{e}_{1s} = \begin{bmatrix}1\\0\end{bmatrix}_s, \mathsf{e}_{2s} = \begin{bmatrix}0\\1\end{bmatrix}_s\Big\}.$$

(a) Given $\mathsf{x}_u = \begin{bmatrix}3\\2\end{bmatrix}_u$ find x_s.

(b) Find e_{1u} and e_{2u}.

5.5.4.- Consider the ordered bases of $\mathbb{R}^2$

$$\mathcal{B} = \Big(\mathsf{b}_{1s} = \begin{bmatrix}1\\2\end{bmatrix}_s, \mathsf{b}_{2s} = \begin{bmatrix}1\\-2\end{bmatrix}_s\Big)$$

$$\mathcal{C} = \Big(\mathsf{c}_{1s} = \begin{bmatrix}1\\1\end{bmatrix}_s, \mathsf{c}_{2s} = \begin{bmatrix}-1\\1\end{bmatrix}_s\Big),$$

where $\mathcal{S}$ is the standard ordered basis.

(a) Given $\mathsf{x}_c = \begin{bmatrix}2\\3\end{bmatrix}_c$, find x_s.

(b) For the same x above, find x_b.

5.5.5.- Let $\mathcal{S} = (\boldsymbol{e}_1, \boldsymbol{e}_2)$ be the standard ordered basis of $\mathbb{R}^2$ and $\mathcal{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ be another ordered basis. Let $\mathsf{A} = \begin{bmatrix}1 & 2\\2 & 3\end{bmatrix}$ be the matrix that transforms the components of a vector $\boldsymbol{x} \in \mathbb{R}^2$ from the basis $\mathcal{S}$ into the basis $\mathcal{B}$, that is, $\mathsf{x}_b = \mathsf{A}\mathsf{x}_s$. Find the components of the basis vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ in the standard basis, that is, find b_{1s} and b_{2s}.

5.5.6.- Show that similarity is a transitive property, that is, if matrix A is similar to matrix B, and B is similar to matrix C, then A is similar to C.

5.5.7.- Consider $\mathbb{R}^3$ with the standard ordered basis $\mathcal{S}$ and the ordered basis

$$\mathcal{U} = \Big(\begin{bmatrix}1\\0\\0\end{bmatrix}_s, \begin{bmatrix}1\\1\\0\end{bmatrix}_s, \begin{bmatrix}1\\1\\1\end{bmatrix}_s\Big).$$

Let $\boldsymbol{T}: \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator

$$\Big[\boldsymbol{T}\Big(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}_s\Big)\Big]_s = \begin{bmatrix}x_1 + 2x_2 - x_3\\-x_2\\x_1 + 7x_3\end{bmatrix}_s.$$

Find both matrices T_{ss} and T_{uu}.

5.5.8.- Consider $\mathbb{R}^2$ with ordered bases

$$\mathcal{S} = \Big(\mathsf{e}_{1s} = \begin{bmatrix}1\\0\end{bmatrix}_s, \mathsf{e}_{2s} = \begin{bmatrix}0\\1\end{bmatrix}_s\Big),$$

$$\mathcal{U} = \Big(\mathsf{u}_{1s} = \begin{bmatrix}1\\1\end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix}1\\-1\end{bmatrix}_s\Big).$$

Let $\boldsymbol{T}: \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation given by

$$[\boldsymbol{T}(\boldsymbol{u}_1)]_s = \begin{bmatrix}1\\3\end{bmatrix}_s, \quad [\boldsymbol{T}(\boldsymbol{u}_2)]_s = \begin{bmatrix}3\\1\end{bmatrix}_s.$$

Find the matrix T_{us}, then the matrices T_{ss}, T_{uu}, and finally T_{su}.

Chapter 6. Inner product spaces

An inner product space is a vector space with an additional structure called inner product. This additional structure is an operation that associates each pair of vectors in the vector space with a scalar. An inner product extends to any vector space the main concepts included in the dot product, which is defined on $\mathbb{R}^n$. These main concepts include the length of a vector, the notioin of perpendicular vectors, and distance between vectors. When these ideas are introduced in function vector spaces, they allow to define the notion of convergence of an infinite sum of vectors. This, in turns, provides a way to evaluate the accuracy of approximate solutions to differential equations.

6.1. Dot product

6.1.1. **Dot product in $\mathbb{R}^2$.** We review the definition of the dot product between vectors in $\mathbb{R}^2$, and we describe its main properties, including the Cauchy-Schwarz inequality. We then use the dot product to introduce the notion of length of a vector, distance and angle between vectors, including the special case of perpendicular vectors. We then review that all these notions can be generalized in a straightforward way from $\mathbb{R}^2$ to $\mathbb{F}^n$, $n \geqslant 1$.

Definition 6.1.1. *Given any vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *with components* $\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ *in the standard ordered basis* $\mathcal{S}$. *The* ***dot product*** *on* $\mathbb{R}^2$ *with is the function* $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

$$\mathsf{x} \cdot \mathsf{y} = x_1 y_1 + x_2 y_2.$$

The ***dot product norm*** *of a vector* $\mathsf{x} \in \mathbb{R}^2$ *is the value of the function* $\| \ \| : \mathbb{R}^2 \to \mathbb{R}$,

$$\|\mathsf{x}\| = \sqrt{\mathsf{x} \cdot \mathsf{x}}.$$

The ***norm distance*** *between* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *is the value of the function* $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

$$d(\mathsf{x}, \mathsf{y}) = \|\mathsf{x} - \mathsf{y}\|.$$

The dot product can be expressed using the transpose of a vector components in the standard basis, as follows,

$$\mathsf{x}^T \mathsf{y} = [x_1, \ x_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1 y_1 + x_2 y_2 = \mathsf{x} \cdot \mathsf{y}.$$

The dot product norm and the norm distance can be expressed in term of vector components in the standard ordered basis $\mathcal{S}$ as follows,

$$\|\mathsf{x}\| = \sqrt{(x_1)^2 + (x_2)^2}, \qquad d(\mathsf{x}, \mathsf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

The geometrical meaning of the norm and distance is clear form this expression in components, as is shown in Fig. 40. The norm of a vector is the Euclidean length from the origin point to the head point of the vector, while the distance between two vectors in the Euclidean distance between the head points of the two vectors.

It is important that we summarize the main properties of the dot product in $\mathbb{R}^2$, since they are the main guide to construct the generalizations of the dot product to other vector spaces.

Theorem 6.1.2. *The dot product on* $\mathbb{R}^2$ *satisfies, for every vector* $\mathsf{x}, \mathsf{y}, \mathsf{z} \in \mathbb{R}^2$ *and every scalar* $a, b \in \mathbb{R}$, *the following properties:*

(a) $\mathsf{x} \cdot \mathsf{y} = \mathsf{y} \cdot \mathsf{x}$, *(Symmetry);*
(b) $\mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z}) = a\,(\mathsf{x} \cdot \mathsf{y}) + b\,(\mathsf{x} \cdot \mathsf{z})$, *(Linearity on the second argument);*
(c) $\mathsf{x} \cdot \mathsf{x} \geqslant 0$, *and* $\mathsf{x} \cdot \mathsf{x} = 0$ *iff* $\mathsf{x} = \mathsf{0}$, *(Positive definiteness).*

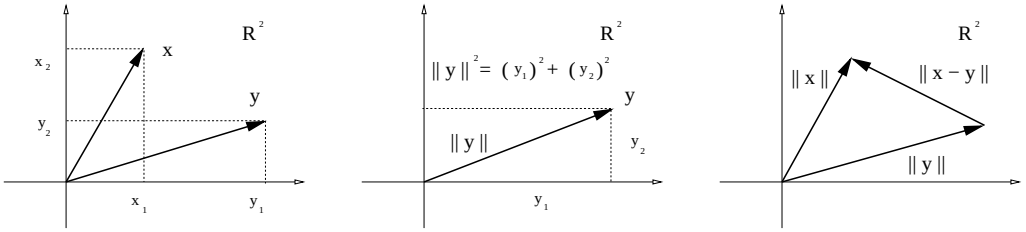

FIGURE 40. Example of the Euclidean notions of vector length and distance between vectors in $\mathbb{R}^2$.

Proof of Theorem 6.1.2: These properties are simple to obtain from the definition of the dot product.

Part (a): It is simple to see that

$$\mathsf{x} \cdot \mathsf{y} = x_1 y_1 + x_2 y_2 = y_1 x_1 + y_2 x_2 = \mathsf{y} \cdot \mathsf{x}.$$

Part (b): It is also simple to see that

$$\begin{aligned} \mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z}) &= x_1(a y_1 + b z_1) + x_2(a y_2 + b z_2) \\ &= a\,(x_1 y_1 + x_2 y_2) + b\,(x_1 z_1 + x_2 z_2) \\ &= a\,(\mathsf{x} \cdot \mathsf{y}) + b\,(\mathsf{x} \cdot \mathsf{z}). \end{aligned}$$

Part (c): This follows from

$$\mathsf{x} \cdot \mathsf{x} = (x_1)^2 + (x_2)^2 \geqslant 0;$$

furthermore, in the case $\mathsf{x} \cdot \mathsf{x} = 0$ we obtain that

$$(x_1)^2 + (x_2)^2 = 0 \quad \Leftrightarrow \quad x_1 = x_2 = 0.$$

This establishes the Theorem. □

These simple properties are crucial to establish the following result, known as Cauchy-Schwarz inequality for the dot product in $\mathbb{R}^2$. This inequality allows to express the dot product of two vectors in $\mathbb{R}^2$ in terms of the angle between the vectors.

Theorem 6.1.3 (Cauchy-Schwarz). *The properties (a)-(c) in Theorem 6.1.2 imply that for all* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *holds*

$$|\mathsf{x} \cdot \mathsf{y}| \leqslant \|\mathsf{x}\|\,\|\mathsf{y}\|.$$

Proof of Theorem 6.1.3: From the positive definiteness property we know that the following inequality holds for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ and for all $a \in \mathbb{R}$,

$$0 \leqslant \|a\mathsf{x} - \mathsf{y}\|^2 = (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}).$$

The symmetry and the linearity on the second argument imply

$$0 \leqslant (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}) = a^2\,\|\mathsf{x}\|^2 - 2a\,(\mathsf{x} \cdot \mathsf{y}) + \|\mathsf{y}\|^2. \tag{6.1}$$

Since the inequality above holds for all $a \in \mathbb{R}$, let us choose a particular value of a, the solution of the equation

$$a\,\|\mathsf{x}\|^2 - (\mathsf{x} \cdot \mathsf{y}) = 0 \quad \Rightarrow \quad a = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}.$$

Introduce this particular value of a into Eq. (6.1),

$$0 \leqslant -\Big(\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}\Big)(\mathsf{x} \cdot \mathsf{y}) + \|\mathsf{y}\|^2 \quad \Rightarrow \quad |\mathsf{x} \cdot \mathsf{y}|^2 \leqslant \|\mathsf{x}\|^2\,\|\mathsf{y}\|^2.$$

This establishes the Theorem. □

The Cauchy-Schwarz inequality implies that we can express the dot product of two vectors in an alternative and more geometrical way, in terms of an angle related with the two vectors. The Cauchy-Schwarz inequality says

$$-1 \leqslant \frac{\mathsf{x}\cdot\mathsf{y}}{\|\mathsf{x}\|\,\|\mathsf{y}\|} \leqslant 1,$$

which suggests that the number $(\mathsf{x}\cdot\mathsf{y})/(\|\mathsf{x}\|\,\|\mathsf{y}\|)$ can be expressed as a sine or a cosine of an appropriate angle.

Theorem 6.1.4. *The **angle** between vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *is the number* $\theta \in [0, \pi]$ *given by*

$$\cos(\theta) = \frac{\mathsf{x}\cdot\mathsf{y}}{\|\mathsf{x}\|\,\|\mathsf{y}\|}.$$

Proof of Theorem 6.1.4: It is not difficult to see that given any vectors $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$, the vectors $\mathsf{x}/\|\mathsf{x}\|$ and $\mathsf{y}/\|\mathsf{y}\|$ have unit norm. Indeed,

$$\left\| \frac{\mathsf{x}}{\|\mathsf{x}\|} \right\|^2 = \frac{(x_1)^2}{\|\mathsf{x}\|^2} + \frac{(x_2)^2}{\|\mathsf{x}\|^2} = \frac{1}{\|\mathsf{x}\|^2}\left[(x_1)^2 + (x_2)^2\right] = 1.$$

The same holds for the vector $\mathsf{y}/\|\mathsf{y}\|$. The expression

$$\frac{\mathsf{x}\cdot\mathsf{y}}{\|\mathsf{x}\|\,\|\mathsf{y}\|} = \frac{\mathsf{x}}{\|\mathsf{x}\|} \cdot \frac{\mathsf{y}}{\|\mathsf{y}\|},$$

shows that the number $(\mathsf{x}\cdot\mathsf{y})/(\|\mathsf{x}\|\,\|\mathsf{y}\|)$ is the inner product of two vectors in the unit circle, as shown in Fig. 41.

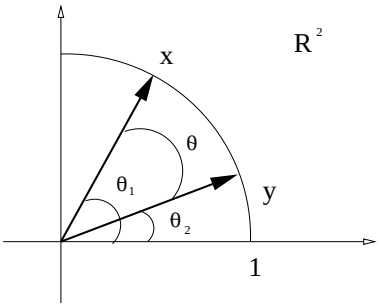

FIGURE 41. The dot product of two vectors $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ can be expressed in terms of the angle $\theta = \theta_1 - \theta_2$ between the vectors.

Therefore, we know that

$$\frac{\mathsf{x}}{\|\mathsf{x}\|} = \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{bmatrix}, \qquad \frac{\mathsf{y}}{\|\mathsf{y}\|} = \begin{bmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{bmatrix}.$$

Their dot product is given by

$$\frac{\mathsf{x}}{\|\mathsf{x}\|} \cdot \frac{\mathsf{y}}{\|\mathsf{y}\|} = \begin{bmatrix} \cos(\theta_1), \sin(\theta_1) \end{bmatrix} \begin{bmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{bmatrix} = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2).$$

Using the formula $\cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2) = \cos(\theta_1 - \theta_2)$, and denoting the angle between the vectors by $\theta = \theta_1 - \theta_2$, we conclude that

$$\frac{\mathsf{x}\cdot\mathsf{y}}{\|\mathsf{x}\|\,\|\mathsf{y}\|} = \cos(\theta).$$

This establishes the Theorem. □

Recall the notion of perpendicular vectors.

Definition 6.1.5. *The vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *are* ***orthogonal****, denoted as* $\mathsf{x} \perp \mathsf{y}$*, iff the angle* $\theta \in [0, \pi]$ *between the vectors is* $\theta = \pi/2$.

The notion of orthogonal vectors in Def. 6.1.5 can be expressed in terms of the dot product, and it is equivalent to Pythagoras Theorem on right triangles.

Theorem 6.1.6. *Let* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *be non-zero vectors, then the following statement holds,*

$$\mathsf{x} \perp \mathsf{y} \quad \Leftrightarrow \quad \mathsf{x} \cdot \mathsf{y} = 0 \quad \Leftrightarrow \quad \|\mathsf{x} - \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2.$$

Proof of Theorem 6.1.6: The non-zero vectors x and $\mathsf{y} \in \mathbb{R}^2$ are orthogonal iff $\theta = \pi/2$, which is equivalent to

$$\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} = 0 \quad \Leftrightarrow \quad \mathsf{x} \cdot \mathsf{y} = 0.$$

The last part of the Proposition comes from the following calculation,

$$\begin{aligned} \|\mathsf{x} - \mathsf{y}\|^2 &= (x_1 - y_1)^2 + (x_2 - y_2)^2 \\ &= (x_1)^2 + (x_2)^2 + (y_1)^2 + (y_2)^2 - 2(x_1 y_1 + x_2 y_2) \\ &= \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2 - 2\,\mathsf{x} \cdot \mathsf{y}. \end{aligned}$$

Hence, $\mathsf{x} \perp \mathsf{y}$ iff $\mathsf{x} \cdot \mathsf{y} = 0$ iff Pythagoras Theorem holds for the triangle with sides given by x, y and hypotenuse $\mathsf{x} - \mathsf{y}$. This establishes the Theorem. □

EXAMPLE 6.1.1: Find the length of the vectors $\mathsf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathsf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, the angle between them, and then find a non-zero vector z orthogonal to x.

SOLUTION: We first find the length, that is, the norms of x and y,

$$\|\mathsf{x}\|^2 = \mathsf{x} \cdot \mathsf{x} = \mathsf{x}^T \mathsf{x} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 + 4 \quad \Rightarrow \quad \|\mathsf{x}\| = \sqrt{5},$$

$$\|\mathsf{y}\|^2 = \mathsf{y} \cdot \mathsf{y} = \mathsf{y}^T \mathsf{y} = \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 9 + 1 \quad \Rightarrow \quad \|\mathsf{y}\| = \sqrt{10}.$$

We now find the angle between x and y,

$$\cos(\theta) = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} = \frac{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix}}{\sqrt{5}\,\sqrt{10}} = \frac{5}{5\sqrt{2}} = \frac{1}{\sqrt{2}} \quad \Rightarrow \quad \theta = \frac{\pi}{4}.$$

We now find z such that $\mathsf{z} \perp \mathsf{x}$, that is,

$$0 = \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = z_1 + 2z_2 \quad \Rightarrow \quad \left\{ \begin{array}{l} z_1 = -2z_2 \\ z_2 \text{ free variable} \end{array} \right\} \quad \Rightarrow \quad \mathsf{z} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} z_2.$$

◁

6.1.2. **Dot product in $\mathbb{F}^n$.** The notion of dot product reviewed above can be generalized in a straightforward way from $\mathbb{R}^2$ to $\mathbb{F}^n$, $n \geqslant 1$, where $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 6.1.7. *The* ***dot product*** *on the vector space* $\mathbb{F}^n$*, with* $n \geqslant 1$*, is the function* $\cdot : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ *given by*

$$\mathsf{x} \cdot \mathsf{y} = \mathsf{x}^* \, \mathsf{y},$$

where x, y *denote components in the standard basis of* $\mathbb{F}^n$*. The* ***dot product norm*** *of a vector* $\mathsf{x} \in \mathbb{F}^n$ *is the value of the function* $\| \ \| : \mathbb{F}^n \to \mathbb{R}$,

$$\|\mathsf{x}\| = \sqrt{\mathsf{x} \cdot \mathsf{x}}.$$

The ***norm distance*** *between* $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ *is the value of the function* $d : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{R}$,

$$d(\mathsf{x}, \mathsf{y}) = \|\mathsf{x} - \mathsf{y}\|.$$

The vectors $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ *are* ***orthogonal****, denoted as* $\mathsf{x} \perp \mathsf{y}$*, iff holds* $\mathsf{x} \cdot \mathsf{y} = 0$.

Notice that we defined two vectors to be orthogonal by the condition that their dot product vanishes. This is the appropriate generalization to $\mathbb{F}^n$ of the ideas we saw in $\mathbb{R}^2$. The concept of angle is more difficult to study. In the case that $\mathbb{F} = \mathbb{C}$ is not clear what the angle between vectors mean. In the case $\mathbb{F} = \mathbb{R}$ and $n > 3$ we have to define angle by the number $(\mathsf{x} \cdot \mathsf{y})/(\|\mathsf{x}\| \, \|\mathsf{y}\|)$. This will be done after we prove the Cauchy-Schwarz inequality, which then is used to show that the number $(\mathsf{x} \cdot \mathsf{y})/(\|\mathsf{x}\| \, \|\mathsf{y}\|) \in [-1, 1]$. The formulas above for the dot product, norm and distance can be expressed in terms of the vector components in the standard basis as follows,

$$\begin{aligned} \mathsf{x} \cdot \mathsf{y} &= \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n, \\ \|\mathsf{x}\| &= \sqrt{|x_1|^2 + \cdots + |x_n|^2}, \\ d(\mathsf{x}, \mathsf{y}) &= \sqrt{|x_1 - y_1|^2 + \cdots + |x_n - y_n|^2}, \end{aligned}$$

where we used the standard notation $\mathsf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\mathsf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, and $|x_i|^2 = \overline{x}_i x_i$, for $i = 1, \cdots n$. In the particular case that $\mathbb{F} = \mathbb{R}$ all the vector components are real numbers, so $\overline{x}_i = x_i$.

EXAMPLE 6.1.2: Find whether x is orthogonal to y and/or z, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad \mathsf{y} = \begin{bmatrix} -5 \\ 4 \\ -3 \\ 2 \end{bmatrix}, \quad \mathsf{z} = \begin{bmatrix} -4 \\ -3 \\ 2 \\ 1 \end{bmatrix}.$$

SOLUTION: We need to compute the dot products $\mathsf{x} \cdot \mathsf{y}$ and $\mathsf{x} \cdot \mathsf{z}$. We obtain

$$\mathsf{x}^T \mathsf{y} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -5 \\ 4 \\ -3 \\ 2 \end{bmatrix} = -5 + 8 - 9 + 8 \quad \Rightarrow \quad \mathsf{x} \cdot \mathsf{y} = 2 \quad \Rightarrow \quad \mathsf{x} \not\perp \mathsf{y},$$

$$\mathsf{x}^T \mathsf{z} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -4 \\ -3 \\ 2 \\ 1 \end{bmatrix} = -4 - 6 + 6 + 4 \quad \Rightarrow \quad \mathsf{x} \cdot \mathsf{z} = 0 \quad \Rightarrow \quad \mathsf{x} \perp \mathsf{z}.$$

◁

EXAMPLE 6.1.3: Find $\mathsf{x} \cdot \mathsf{y}$, where $\mathsf{x} = \begin{bmatrix} 2 + 3i \\ i \\ 1 - i \end{bmatrix}$ and $\mathsf{y} = \begin{bmatrix} 2i \\ 1 \\ 1 + 3i \end{bmatrix}$.

SOLUTION: The first product $\mathsf{x} \cdot \mathsf{y}$ is given by

$$\mathsf{x}^* \mathsf{y} = \begin{bmatrix} 2 - 3i & -i & 1 + i \end{bmatrix} \begin{bmatrix} 2i \\ 1 \\ 1 + 3i \end{bmatrix} = (2 - 3i)(2i) - i + (1 + i)(1 + 3i),$$

so $\mathsf{x} \cdot \mathsf{y} = 4i + 6 - i + 1 - 3 + i + 3i$, that is, $\mathsf{x} \cdot \mathsf{y} = 4 + 7i$. ◁

The dot product satisfies the following properties.

Theorem 6.1.8. *The dot product on $\mathbb{F}^n$, with $n \geqslant 1$, satisfies for every vector $\mathsf{x}, \mathsf{y}, \mathsf{z} \in \mathbb{F}^n$ and every scalar $a, b \in \mathbb{F}$, the following properties:*

(a1) $\mathsf{x}\cdot\mathsf{y} = \mathsf{y}\cdot\mathsf{x}$, *(Symmetry $\mathbb{F} = \mathbb{R}$);*

(a2) $\mathsf{x}\cdot\mathsf{y} = \overline{\mathsf{y}\cdot\mathsf{x}}$, *(Conjugate symmetry, for $\mathbb{F} = \mathbb{C}$);*

(b) $\mathsf{x}\cdot(a\mathsf{y} + b\mathsf{z}) = a\,(\mathsf{x}\cdot\mathsf{y}) + b\,(\mathsf{x}\cdot\mathsf{z})$, *(Linearity on the second argument);*

(c) $\mathsf{x}\cdot\mathsf{x} \geqslant 0$, *and* $\mathsf{x}\cdot\mathsf{x} = 0$ *iff* $\mathsf{x} = 0$, *(Positive definiteness).*

Proof of Theorem 6.1.8: Use the expression of the dot product in terms of the vector components. The property in (a1) can be established as follows,

$$\mathsf{x}\cdot\mathsf{y} = x_1y_1 + \cdots + x_ny_n = y_1x_1 + \cdots + y_nx_n = \mathsf{y}\cdot\mathsf{x}.$$

The property in (a2) can be established as follows,

$$\mathsf{x}\cdot\mathsf{y} = \overline{x}_1y_1 + \cdots + \overline{x}_ny_n = \overline{(\overline{y}_1x_1 + \cdots + \overline{y}_nx_n)} = \overline{\mathsf{y}\cdot\mathsf{x}}.$$

The property in (b) is shown in a similar way,

$$\mathsf{x}\cdot(a\mathsf{y} + b\mathsf{z}) = \mathsf{x}^*(a\,\mathsf{y} + b\,\mathsf{z}) = a\,\mathsf{x}^*\,\mathsf{y} + b\,\mathsf{x}^*\,\mathsf{z} = a\,(\boldsymbol{x}\cdot\boldsymbol{y}) + b\,(\boldsymbol{x}\cdot\boldsymbol{z}).$$

The property in (c) follows from

$$\mathsf{x}\cdot\mathsf{x} = \mathsf{x}^*\,\mathsf{x} = |x_1|^2 + \cdots + |x_n|^2 \geqslant 0;$$

furthermore, in the case that $\mathsf{x}\cdot\mathsf{x} = 0$ we obtain that

$$|x_1|^2 + \cdots + |x_n|^2 = 0 \quad\Leftrightarrow\quad x_1 = \cdots = x_n = 0 \quad\Leftrightarrow\quad \mathsf{x} = 0.$$

This establishes the Theorem. □

The positive definiteness property (c) above shows that the dot product norm is indeed a real-valued and not a complex-valued function, since $\mathsf{x}\cdot\mathsf{x} \geqslant 0$ implies that $\|\mathsf{x}\| = \sqrt{\mathsf{x}\cdot\mathsf{x}} \in \mathbb{R}$. In the case of $\mathbb{F} = \mathbb{R}$, the symmetry property and the linearity in the second argument property imply that the dot product on $\mathbb{R}^n$ is also linear in the first argument. This is a reason to call the dot product on $\mathbb{R}^n$ a bilinear form. Finally, notice that in the case $\mathbb{F} = \mathbb{C}$, the conjugate symmetry property and the linearity in the second argument imply that *the dot product on $\mathbb{C}^n$ is conjugate linear on the first argument*. The proof is the following:

$$(a\mathsf{y} + b\mathsf{z})\cdot\mathsf{x} = \overline{\mathsf{x}\cdot(a\mathsf{y} + b\mathsf{z})} = \overline{a\,(\mathsf{x}\cdot\mathsf{y}) + b\,(\mathsf{x}\cdot\mathsf{z})} = \overline{a}\,\overline{(\mathsf{x}\cdot\mathsf{y})} + \overline{b}\,\overline{(\mathsf{x}\cdot\mathsf{z})},$$

that is, for all $\mathsf{x}, \mathsf{y}, \mathsf{z} \in \mathbb{C}^n$ and all $a, b \in \mathbb{C}$ holds

$$(a\mathsf{y} + b\mathsf{z})\cdot\mathsf{x} = \overline{a}\,(\mathsf{y}\cdot\mathsf{x}) + \overline{b}\,(\mathsf{z}\cdot\mathsf{x}).$$

Hence we say that the dot product on $\mathbb{C}^n$ is conjugate linear in the first argument.

EXAMPLE 6.1.4: Compute the dot product of $\mathsf{x} = \begin{bmatrix} 2+3i \\ 6i-9 \end{bmatrix}$ with $\mathsf{y} = \begin{bmatrix} 3i \\ 2 \end{bmatrix}$.

SOLUTION: This is a straightforward computation

$$\mathsf{x}\cdot\mathsf{y} = \mathsf{x}^*\mathsf{y} = \begin{bmatrix} 2-3i & -6i-9 \end{bmatrix}\begin{bmatrix} 3i \\ 2 \end{bmatrix} = 6i + 9 - 12i - 18 \quad\Rightarrow\quad \mathsf{x}\cdot\mathsf{y} = -9 - 6i.$$

Notice that $\mathsf{x} = (2+3i)\hat{\mathsf{x}}$, with $\hat{\mathsf{x}} = \begin{bmatrix} 1 \\ 3i \end{bmatrix}$, so we could use the conjugate linearity in the first argument to compute

$$\mathsf{x}\cdot\mathsf{y} = \big((2+3i)\hat{\mathsf{x}}\big)\cdot\mathsf{y} = (2-3i)\,(\hat{\mathsf{x}}\cdot\mathsf{y}) = (2-3i)\begin{bmatrix} 1 & -3i \end{bmatrix}\begin{bmatrix} 3i \\ 2 \end{bmatrix} = (2-3i)(3i-6i),$$

and we obtain the same result, $\mathsf{x}\cdot\mathsf{y} = -9 - 6i$. Finally, notice that $\mathsf{y}\cdot\mathsf{x} = -9 + 6i$. ◁

An important result is that the dot product in $\mathbb{F}^n$ satisfies the Cauchy-Schwarz inequality.

Theorem 6.1.9 (Cauchy-Schwarz). *The properties (a1)-(c) in Theorem 6.1.8 imply that for all* $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ *holds*

$$|\mathsf{x} \cdot \mathsf{y}| \leqslant \|\mathsf{x}\| \, \|\mathsf{y}\|.$$

REMARK: The proof of the Cauchy-Schwarz inequality only uses the three properties of the dot product presented in Theorem 6.1.8. Any other function $f : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ having these three properties also satisfies the Cauchy-Schwarz inequality.

Proof of Theorem 6.1.9: From the positive definiteness property we know that the following inequality holds for all $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ and for all $a \in \mathbb{F}$,

$$0 \leqslant \|a\mathsf{x} - \mathsf{y}\|^2 = (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}).$$

The symmetry and the linearity on the second argument imply

$$0 \leqslant (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}) = a\,\overline{a}\,\|\mathsf{x}\|^2 - \overline{a}\,(\mathsf{x} \cdot \mathsf{y}) - a\,(\mathsf{y} \cdot \mathsf{x}) + \|\mathsf{y}\|^2. \tag{6.2}$$

Since the inequality above holds for all $a \in \mathbb{F}$, let us choose a particular value of a, the solution of the equation

$$a\,\overline{a}\,\|\mathsf{x}\|^2 - \overline{a}\,(\mathsf{x} \cdot \mathsf{y}) = 0 \quad \Rightarrow \quad a = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}.$$

Introduce this particular value of a into Eq. (6.2),

$$0 \leqslant -\Big(\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}\Big)(\overline{\mathsf{x} \cdot \mathsf{y}}) + \|\mathsf{y}\|^2 \quad \Rightarrow \quad |\mathsf{x} \cdot \mathsf{y}|^2 \leqslant \|\mathsf{x}\|^2 \, \|\mathsf{y}\|^2.$$

This establishes the Theorem. □

In the case $\mathbb{F} = \mathbb{R}$, the Cauchy-Schwarz inequality in $\mathbb{R}^n$ implies that the number $(\mathsf{x} \cdot \mathsf{y})/(\|\mathsf{x}\| \, \|\mathsf{y}\|) \in [-1, 1]$, which is a necessary and sufficient condition for the following definition of angle between two vectors in $\mathbb{R}^n$.

Definition 6.1.10. *The* ***angle*** *between vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^n$ *is the number* $\theta \in [0, \pi]$ *given by*

$$\cos(\theta) = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|}.$$

The dot product norm function in Definition 6.1.7 satisfies the following properties.

Theorem 6.1.11. *The dot product norm function on* $\mathbb{F}^n$*, with* $n \geqslant 1$*, satisfies for every vector* $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ *and every scalar* $a \in \mathbb{F}$ *the following properties:*

(a) $\|\mathsf{x}\| \geqslant 0$*, and* $\|\mathsf{x}\| = 0$ *iff* $\mathsf{x} = 0$*, (Positive definiteness);*
(b) $\|a\mathsf{x}\| = |a| \, \|\mathsf{x}\|$*, (Scaling);*
(c) $\|\mathsf{x} + \mathsf{y}\| \leqslant \|\mathsf{x}\| + \|\mathsf{y}\|$*, (Triangle inequality).*

Proof of Theorem 6.1.11: Properties (a) and (b) are straightforward to show from the definition of dot product, and their proof is left as an exercise. We show here how to obtain the triangle inequality, property (c). The proof uses the Cauchy-Schwarz inequality presented in Theorem 6.1.9. Given any vectors $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ holds

$$\begin{aligned} \|\mathsf{x} + \mathsf{y}\|^2 &= (\mathsf{x} + \mathsf{y}) \cdot (\mathsf{x} + \mathsf{y}) \\ &= \|\mathsf{x}\|^2 + (\mathsf{x} \cdot \mathsf{y}) + (\mathsf{y} \cdot \mathsf{x}) + \|\mathsf{y}\|^2 \\ &\leqslant \|\mathsf{x}\|^2 + 2\,|\mathsf{x} \cdot \mathsf{y}| + \|\mathsf{y}\|^2 \\ &\leqslant \|\mathsf{x}\|^2 + 2\,\|\mathsf{x}\| \, \|\mathsf{y}\| + \|\mathsf{y}\|^2 = \big(\|\mathsf{x}\| + \|\mathsf{y}\|\big)^2, \end{aligned}$$

We conclude that $\|\mathsf{x} + \mathsf{y}\|^2 \leqslant \big(\|\mathsf{x}\| + \|\mathsf{y}\|\big)^2$. This establishes the Theorem. □

A vector $\mathsf{v} \in \mathbb{F}^n$ is called *normal or unit vector* iff $\|\mathsf{v}\| = 1$. Examples of unit vectors are the standard basis vectors. Unit vectors parallel to a given vector are simple to find.

Theorem 6.1.12. *If* $\mathsf{v} \in \mathbb{F}^n$ *is non-zero, then* $\dfrac{\mathsf{v}}{\|\mathsf{v}\|}$ *is a unit vector parallel to* v.

Proof of Theorem 6.1.12: Notice that $\mathsf{u} = \dfrac{\mathsf{v}}{\|\mathsf{v}\|}$ is parallel to v, and it is straightforward to check that u is a unit vector, since

$$\|\mathsf{u}\| = \left\| \frac{\mathsf{v}}{\|\mathsf{v}\|} \right\| = \frac{1}{\|\mathsf{v}\|} \|\mathsf{v}\| = 1.$$

This establishes the Theorem. □

EXAMPLE 6.1.5: Find a unit vector parallel to $\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

SOLUTION: First compute the norm of x,

$$\|\mathsf{x}\| = \sqrt{1+4+9} = \sqrt{14},$$

therefore $\mathsf{u} = \dfrac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is a unit vector parallel to v. ◁

6.1.3. **Exercises.**

6.1.1.- Consider the vector space $\mathbb{R}^4$ with standard basis $\mathcal{S}$ and dot product. Find the norm of u and v, their distance and the angle between them, where

$$\mathsf{u} = \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

6.1.2.- Use the dot product on $\mathbb{R}^2$ to find two unit vectors orthogonal to

$$\mathsf{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

6.1.3.- Use the dot product on $\mathbb{C}^2$ to find a unit vector parallel to

$$\mathsf{x} = \begin{bmatrix} 1+2i \\ 2-i \end{bmatrix}.$$

6.1.4.- Consider the vector space $\mathbb{R}^2$ with the dot product.

(a) Give an example of a linearly independent set $\{\mathsf{x}, \mathsf{y}\}$ with $\mathsf{x} \not\perp \mathsf{y}$.

(b) Give an example of a linearly dependent set $\{\mathsf{x}, \mathsf{y}\}$ with $\mathsf{x} \perp \mathsf{y}$.

6.1.5.- Consider the vector space $\mathbb{F}^n$ with the dot product, and let Re denote the real part of a complex number. Show that for all x, $\mathsf{y} \in \mathbb{F}^n$ holds

$$\|\mathsf{x} - \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2 - 2\mathsf{Re}(\mathsf{x} \cdot \mathsf{y}).$$

6.1.6.- Use the result in Exercise **6.1.5** above to prove the following generalizations of the Pythagoras Theorem to $\mathbb{F}^n$ with the dot product.

(a) For x, $\mathsf{y} \in \mathbb{R}^n$ holds

$$\mathsf{x} \perp \mathsf{y} \quad \Leftrightarrow \quad \|\mathsf{x} + \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2.$$

(b) For x, $\mathsf{y} \in \mathbb{C}^n$ holds

$$\mathsf{x} \perp \mathsf{y} \quad \Rightarrow \quad \|\mathsf{x} + \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2.$$

6.1.7.- Prove that the parallelogram law holds for the dot product norm in $\mathbb{F}^n$, that is, show that for all x, $\mathsf{y} \in \mathbb{F}^n$ holds

$$\|\mathsf{x} + \mathsf{y}\|^2 + \|\mathsf{x} - \mathsf{y}\|^2 = 2\|\mathsf{x}\|^2 + 2\|\mathsf{y}\|^2.$$

This law states that the sum of the squares of the lengths of the four sides of a parallelogram formed by x and y equals the sum of the square of the lengths of the two diagonals.

6.2. Inner product

6.2.1. **Inner product.** An inner product on a vector space is a generalization of the dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$ introduced in Sect. 6.1. The inner product is not defined with a particular formula, or requiring a particular basis in the vector space. Instead, the inner product is defined by a list of properties that must satisfy. We did something similar when we introduced the concept of a vector space. In that case we defined a vector space as a set of any kind of elements where linear combinations are possible, instead of defining the set by explicitly giving its elements.

Definition 6.2.1. *Let V be a vector space over the scalar field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A function $\langle\,,\rangle : V \times V \to \mathbb{F}$ is called an* ***inner product*** *iff for every $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z} \in V$ and every a, $b \in \mathbb{F}$ the function $\langle\,,\rangle$ satisfies:*

(a1) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$, *(Symmetry, for $\mathbb{F} = \mathbb{R}$);*

(a2) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$, *(Conjugate symmetry, for $\mathbb{F} = \mathbb{C}$);*

(b) $\langle \boldsymbol{x}, (a\boldsymbol{y} + b\boldsymbol{z})\rangle = a\langle \boldsymbol{x}, \boldsymbol{y}\rangle + b\langle \boldsymbol{x}, \boldsymbol{z}\rangle$, *(Linearity on the second argument);*

(c) $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geqslant 0$, *and* $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ *iff* $\boldsymbol{x} = 0$, *(Positive definiteness).*

An ***inner product space*** *is a pair $\big(V, \langle\,,\rangle\big)$ of a vector space with an inner product.*

Different inner products can be defined on a given vector space. The dot product is an inner product in $\mathbb{F}^n$. A different inner product can be defined in $\mathbb{F}^n$, as can be seen in the following example.

Example 6.2.1: We show that $\mathbb{R}^n$ can have different inner products.

(a) The dot product on $\mathbb{R}^n$ is an inner product, since the expression

$$\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_s^T \mathsf{y}_s = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}_s \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_s = x_1 y_1 + \cdots + x_n y_n, \tag{6.3}$$

satisfies all the properties in Definition 6.2.1, with $\mathcal{S}$ the standard ordered basis in $\mathbb{R}^n$.

(b) A different inner product in $\mathbb{R}^n$ can be introduced by a formula similar to the one in Eq. (6.3) by choosing a different ordered basis. If $\mathcal{U}$ is any ordered basis of $\mathbb{R}^n$, then

$$\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_u^T \mathsf{y}_u.$$

defines an inner product on $\mathbb{R}^n$. The inner product defined using the basis $\mathcal{U}$ is not equal to the inner product defined using the standard basis $\mathcal{S}$. Let $\mathsf{P} = \mathsf{I}_{us}$ be the change of basis matrix, then we know that $\mathsf{x}_u = \mathsf{P}^{-1}\mathsf{x}_s$. The inner product above can be expressed in terms of the $\mathcal{S}$ basis as follows,

$$\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_s^T \mathsf{M}\, \mathsf{y}_s, \qquad \mathsf{M} = \big(\mathsf{P}^{-1}\big)^T \big(\mathsf{P}^{-1}\big),$$

and in general, $\mathsf{M} \neq I_n$. Therefore, the inner product above is not equal to the dot product. Also, see Example 6.2.2.

◁

Example 6.2.2: Let $\mathcal{S}$ be the standard ordered basis in $\mathbb{R}^2$, and introduce the ordered basis $\mathcal{U}$ as the following rescaling of $\mathcal{S}$,

$$\mathcal{U} = \big(\boldsymbol{u}_1 = \frac{1}{2}\,\boldsymbol{e}_1, \boldsymbol{u}_2 = \frac{1}{3}\,\boldsymbol{e}_2\big).$$

Express the inner product $\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_u^T \mathsf{y}_u$ in terms of x_s and y_s. Is this inner product the same as the dot product?

SOLUTION: The definition of the inner product says that $\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_u^T \mathsf{y}_u$. Introducing the notation $\mathsf{x}_u = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}_u$ and $\mathsf{y}_u = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix}_u$, we obtain the usual expression $\langle \mathsf{x}, \mathsf{y} \rangle = \tilde{x}_1 \tilde{y}_1 + \tilde{x}_2 \tilde{y}_2$. The components x_u and x_s are related by the change of basis formula

$$\mathsf{x}_u = \mathsf{P}^{-1} \mathsf{x}_s, \quad \mathsf{P} = \mathsf{I}_{us} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}_{us} \quad \Rightarrow \quad \mathsf{P}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}_{su} = \left(\mathsf{P}^{-1}\right)^T.$$

Therefore,

$$\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_u^T \mathsf{y}_u = \mathsf{x}_s^T \left(\mathsf{P}^{-1}\right)^T \mathsf{P}^{-1} \mathsf{y}_s = \begin{bmatrix} x_1, & x_2 \end{bmatrix}_s \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}_{su} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_s$$

where we used the standard notation $\mathsf{x}_s = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s$ and $\mathsf{y}_s = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_s$. We conclude that

$$\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_s^T \left(\mathsf{P}^{-1}\right)^2 \mathsf{y}_s \quad \Leftrightarrow \quad \langle \mathsf{x}, \mathsf{y} \rangle = 4x_1 y_1 + 9x_2 y_2.$$

The inner product $\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_u^T \mathsf{y}_u$ is different from the dot product $\mathsf{x} \cdot \mathsf{y} = \mathsf{x}_s^T \mathsf{y}_s$. ◁

EXAMPLE 6.2.3: Determine whether the function $\langle \, , \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ below is an inner product in $\mathbb{R}^3$, where

$$\langle \mathsf{x}, \mathsf{y} \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3 + 3x_1 y_2 + 3x_2 y_1.$$

SOLUTION: The function $\langle \, , \rangle$ seems to be symmetric and linear. It is not so clear whether this function is positive, because of the presence of crossed terms. So, before spending time to prove the symmetry and linearity properties, we first concentrate on the property that might fail, positivity. If positivity fails, we don't need to prove the remaining properties. The crossed terms $3(x_1 y_2 + x_2 y_1)$ in the definition of the inner product suggest that the product is not positive, since the factor 3 makes them too important compared with the other terms. Let us try to find an example, that is a vector $\mathsf{x} \neq \mathsf{0}$ such that $\langle \mathsf{x}, \mathsf{x} \rangle \leqslant 0$. Let us try with

$$\mathsf{x} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \langle \mathsf{x}, \mathsf{x} \rangle = 1^2 + (-1)^2 + 0^2 + 3[(1)(-1) + (-1)(1)] = 2 - 6 = -4.$$

Since $\langle \mathsf{x}, \mathsf{x} \rangle = -4$, the function $\langle \, , \rangle$ is not positive, hence it is not an inner product. ◁

EXAMPLE 6.2.4: Consider the vector space $\mathbb{F}^{m,n}$ of all $m \times n$ matrices. Show that an inner product on that space is the function $\langle \, , \rangle_F : \mathbb{F}^{m,n} \times \mathbb{F}^{m,n} \to \mathbb{F}$

$$\langle \mathsf{A}, \mathsf{B} \rangle_F = \operatorname{tr}(\mathsf{A}^* \mathsf{B}).$$

The inner product is called *Frobenius inner product*.

SOLUTION: We show that the Frobenius function $\langle \, , \rangle_F$ above satisfies the three properties in Def. 6.2.1. We use the component notation $\mathsf{A} = [A_{ij}]$, $\mathsf{B} = [B_{kl}]$, with $i, k = 1, \cdots, m$ and $j, l = 1, \cdots, n$, so

$$(\mathsf{A}^* \mathsf{B})_{jl} = \sum_{i=1}^{m} \left(\overline{A}^T\right)_{ji} B_{il} = \sum_{i=1}^{m} \overline{A}_{ij} B_{il} \quad \Rightarrow \quad \langle \mathsf{A}, \mathsf{B} \rangle_F = \sum_{j=1}^{n} \sum_{i=1}^{n} \overline{A}_{ij} B_{ij}.$$

The first property is satisfied, since

$$\langle \mathsf{A}, \mathsf{A} \rangle_F = \operatorname{tr}(\mathsf{A}^* \mathsf{A}) = \sum_{j=1}^{n} \sum_{i=1}^{m} |A_{ij}|^2 \geqslant 0,$$

and $\langle \mathsf{A}, \mathsf{A} \rangle_F = 0$ iff $A_{ij} = 0$ for every indices i, j, which is equivalently to $\mathsf{A} = \mathsf{0}$. The second property is satisfied, since

$$\langle \mathsf{A}, \mathsf{B} \rangle_F = \sum_{j=1}^{n} \sum_{i=1}^{n} \overline{A}_{ij} B_{ij} = \overline{\sum_{j=1}^{n} \sum_{i=1}^{n} \overline{B}_{ij} A_{ij}} = \overline{\langle \mathsf{B}, \mathsf{A} \rangle}_F.$$

The same proof can be expressed in index-free notation using the properties of the trace,

$$\operatorname{tr}\left(\mathsf{A}^* \mathsf{B}\right)) = \operatorname{tr}\left(\overline{\mathsf{A}}^T \mathsf{B}\right) = \operatorname{tr}\left[(\overline{\mathsf{A}}^T \mathsf{B})^T\right] = \operatorname{tr}\left(\mathsf{B}^T \overline{\mathsf{A}}\right) = \overline{\operatorname{tr}\left(\mathsf{B}^* \mathsf{A}\right)},$$

that is, $\langle \mathsf{A}, \mathsf{B} \rangle_F = \overline{\langle \mathsf{B}, \mathsf{A} \rangle}_F$. The third property comes from the distributive property of the matrix product, that is,

$$\langle \mathsf{A}, (a\mathsf{B} + b\mathsf{C}) \rangle_F = \operatorname{tr}\left(\mathsf{A}^* \left(a\mathsf{B} + b\mathsf{C}\right)\right) = a \operatorname{tr}\left(\mathsf{A}^* \mathsf{B}\right) + b \operatorname{tr}\left(\mathsf{A}^* \mathsf{C}\right) = a \langle \mathsf{A}, \mathsf{B} \rangle_F + b \langle \mathsf{A}, \mathsf{C} \rangle_F.$$

This establishes that $\langle \, , \rangle_F$ is an inner product. ◁

EXAMPLE 6.2.5: Compute the Frobenius inner product $\langle \mathsf{A}, \mathsf{B} \rangle_F$ where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \in \mathbb{R}^{2,3}.$$

SOLUTION: Since the matrices have real coefficients, the Frobenius inner product has the form $\langle \mathsf{A}, \mathsf{B} \rangle_F = \operatorname{tr}\left(\mathsf{A}^T \mathsf{B}\right)$. So, we need to compute the diagonal elements in the product

$$\mathsf{A}^T \mathsf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & * & * \\ * & 8 & * \\ * & * & 5 \end{bmatrix} \quad \Rightarrow \quad \langle \mathsf{A}, \mathsf{B} \rangle_F = 7 + 8 + 5 \quad \Rightarrow \quad \langle \mathsf{A}, \mathsf{B} \rangle_F = 20.$$

◁

EXAMPLE 6.2.6: Consider the vector space $\mathbb{P}_n([-1, 1])$ of polynomials with real coefficients having degree less or equal $n \geq 1$ and being defined on the interval $[-1, 1]$. Show that an inner product in this space is the following:

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} \boldsymbol{p}(x) \boldsymbol{q}(x) \, dx. \quad \boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_n.$$

SOLUTION: We need to verify the three properties in the Definition 6.2.1. The positive definiteness property is satisfied, since

$$\langle \boldsymbol{p}, \boldsymbol{p} \rangle = \int_{-1}^{1} \left[\boldsymbol{p}(x)\right]^2 dx \geqslant 0,$$

and in the case $\langle \boldsymbol{p}, \boldsymbol{p} \rangle = 0$ this implies that the integrand must vanish, that is, $[\boldsymbol{p}(x)]^2 = 0$, which is equivalent to $\boldsymbol{p} = 0$. The symmetry property is satisfied, since $\boldsymbol{p}(x)\boldsymbol{q}(x) = \boldsymbol{q}(x)\boldsymbol{p}(x)$, which implies that $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \langle \boldsymbol{q}, \boldsymbol{p} \rangle$. The linearity property on the second argument is also satisfied, since

$$\begin{aligned} \langle \boldsymbol{p}, (a\boldsymbol{q} + b\boldsymbol{r}) \rangle &= \int_{-1}^{1} \boldsymbol{p}(x) \left[a \, \boldsymbol{q}(x) + b \, \boldsymbol{r}(x)\right] dx \\ &= a \int_{-1}^{1} \boldsymbol{p}(x) \boldsymbol{q}(x) \, dx + b \int_{-1}^{1} \boldsymbol{p}(x) \boldsymbol{r}(x) \, dx \\ &= a \langle \boldsymbol{p}, \boldsymbol{q} \rangle + b \langle \boldsymbol{p}, \boldsymbol{r} \rangle. \end{aligned}$$

This establishes that $\langle \, , \rangle$ is an inner product. ◁

EXAMPLE 6.2.7: Consider the vector space $C^k([a,b],\mathbb{R})$, with $k \geqslant 0$ and $a < b$, of k-times continuously differentiable real-valued functions $\boldsymbol{f} : [a,b] \to \mathbb{R}$. An inner product in this vector space is given by

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_a^b \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx.$$

Any positive function $\mu \in C^0([a,b],\mathbb{R})$ determines an inner product in $C^k([a,b],\mathbb{R})$ as follows

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle_\mu = \int_a^b \mu(x)\, \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx.$$

The function μ is called a weigh function. An inner product in the vector space $C^k([a,b],\mathbb{C})$ of k-times continuously differentiable complex-valued functions $\boldsymbol{f} : [a,b] \subset \mathbb{R} \to \mathbb{C}$ is the following,

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_a^b \overline{\boldsymbol{f}(x)}\boldsymbol{g}(x)\, dx.$$

◁

An inner product satisfies the following inequality.

Theorem 6.2.2 (Cauchy-Schwarz). *If $\big(V, \langle\,,\,\rangle\big)$ is an inner product space over $\mathbb{F}$, then for every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds*

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x} \rangle\, \langle \boldsymbol{y}, \boldsymbol{y} \rangle.$$

Furthermore, equality holds iff $\boldsymbol{y} = a\boldsymbol{x}$, with $a = \langle \boldsymbol{x}, \boldsymbol{y} \rangle / \langle \boldsymbol{x}, \boldsymbol{x} \rangle$.

Proof of Theorem 6.2.2: From the positive definiteness property we know that for every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and every scalar $a \in \mathbb{F}$ holds $0 \leqslant \langle (a\boldsymbol{x} - \boldsymbol{y}), (a\boldsymbol{x} - \boldsymbol{y}) \rangle$. The symmetry and the linearity on the second argument imply

$$0 \leqslant \langle (a\boldsymbol{x} - \boldsymbol{y}), (a\boldsymbol{x} - \boldsymbol{y}) \rangle = a\,\overline{a}\,\langle \boldsymbol{x}, \boldsymbol{x} \rangle - \overline{a}\,\langle \boldsymbol{x}, \boldsymbol{y} \rangle - a\,\langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle. \tag{6.4}$$

Since the inequality above holds for all $a \in \mathbb{F}$, let us choose a particular value of a, the solution of the equation

$$a\,\overline{a}\,\langle \boldsymbol{x}, \boldsymbol{x} \rangle - \overline{a}\,\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \quad \Rightarrow \quad a = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

Introduce this particular value of a into Eq. (6.4),

$$0 \leqslant -\Big(\frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}\Big)\overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle} + \langle \boldsymbol{y}, \boldsymbol{y} \rangle \quad \Rightarrow \quad |\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x} \rangle\, \langle \boldsymbol{y}, \boldsymbol{y} \rangle.$$

Finally, notice that equality holds iff $a\boldsymbol{x} = \boldsymbol{y}$, and in this case, computing the inner product with $\boldsymbol{x}$ we obtain $a\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. This establishes the Theorem. □

6.2.2. **Inner product norm.** The inner product on a vector space determines a particular notion of length, or norm, of a vector, and we call it the inner product norm. After we introduce this norm we show its main properties. In Chapter 8 later on we use these properties to define a more general notion of norm as any function on the vector space satisfying these properties. The inner product norm is just a particular case of this broader notion of length. A normed space is a vector space with any norm.

Definition 6.2.3. *The **inner product norm** determined in an inner product space $\big(V, \langle\,,\,\rangle\big)$ is the function $\|\ \| : V \to \mathbb{R}$ given by*

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

The Cauchy-Schwarz inequality is often expressed using the inner product norm as follows: For every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|.$$

A vector $\boldsymbol{x} \in V$ is a *normal or unit vector* iff $\|\boldsymbol{x}\| = 1$.

Theorem 6.2.4. *If $\boldsymbol{v} \neq \mathbf{0}$ belongs to $\big(V, \langle\,,\rangle\big)$, then $\dfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ is a unit vector parallel to $\boldsymbol{v}$.*

The proof is the same of Theorem 6.1.12.

EXAMPLE 6.2.8: Consider the inner product space $\big(\mathbb{F}^{m,n}, \langle\,,\rangle_F\big)$, where $\mathbb{F}^{m,n}$ is the vector space of all $m \times n$ matrices and $\langle\,,\rangle_F$ is the Frobenius inner product defined in Example 6.2.4. The associated inner product norm is called the *Frobenius norm* and is given by

$$\|\mathsf{A}\|_F = \sqrt{\langle \mathsf{A}, \mathsf{A} \rangle_F} = \sqrt{\mathrm{tr}\,(\mathsf{A}^* \mathsf{A})}.$$

If $\mathsf{A} = [A_{ij}]$, with $i = 1, \cdots, m$ and $j = 1, \cdots, n$, then

$$\|\mathsf{A}\|_F = \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2 \Big)^{1/2}.$$

◁

EXAMPLE 6.2.9: Find an explicit expression for the Frobenius norm of any element $\mathsf{A} \in \mathbb{F}^{2,2}$.

SOLUTION: The Frobenius norm of an arbitrary matrix $\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathbb{F}^{2,2}$ is given by

$$\|\mathsf{A}\|_F^2 = \mathrm{tr}\, \Big(\begin{bmatrix} \overline{A}_{11} & \overline{A}_{21} \\ \overline{A}_{12} & \overline{A}_{22} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \Big).$$

Since we are only interested in the diagonal elements of the matrix product in the equation above, we obtain

$$\|\mathsf{A}\|_F^2 = \mathrm{tr} \begin{bmatrix} |A_{11}|^2 + |A_{21}|^2 & * \\ * & |A_{12}|^2 + |A_{22}|^2 \end{bmatrix}$$

which gives the formula

$$\|\mathsf{A}\|_F^2 = |A_{11}|^2 + |A_{12}|^2 + |A_{21}|^2 + |A_{22}|^2.$$

This is the explicit expression of the sum $\|\mathsf{A}\|_F = \Big(\sum_{i=1}^{2} \sum_{j=1}^{2} |A_{ij}|^2 \Big)^{1/2}$. ◁

The inner product norm function has the following properties.

Theorem 6.2.5. *The inner product norm introduced in Definition 6.2.3 satisfies that for every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and every $a \in \mathbb{F}$ holds,*

(a) $\|\boldsymbol{x}\| \geqslant 0$, *and* $\|\boldsymbol{x}\| = 0$ *iff* $\boldsymbol{x} = 0$, *(Positive definiteness);*
(b) $\|a\boldsymbol{x}\| = |a| \, \|\boldsymbol{x}\|$, *(Scaling);*
(c) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$, *(Triangle inequality).*

Proof of Theorem 6.2.5: Properties (a) and (b) are straightforward to show from the definition of inner product, and their proof is left as an exercise. We show here how to

obtain the triangle inequality, property (c). Given any vectors $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds

$$\begin{aligned} \|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle (\boldsymbol{x}+\boldsymbol{y}), (\boldsymbol{x}+\boldsymbol{y}) \rangle \\ &= \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \|\boldsymbol{y}\|^2 \\ &\leqslant \|\boldsymbol{x}\|^2 + 2\,|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| + \|\boldsymbol{y}\|^2 \\ &\leqslant \|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 = \big(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|\big)^2, \end{aligned}$$

where the last inequality comes from the Cauchy-Schwarz inequality. We then conclude that $\|\boldsymbol{x}+\boldsymbol{y}\|^2 \leqslant \big(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|\big)^2$. This establishes the Theorem. □

6.2.3. **Norm distance.** The norm on an inner product space determines a particular notion of distance between vectors. After we introduce this norm we show its main properties.

Definition 6.2.6. *The* ***norm distance*** *between two vectors in a vector space V with a norm function $\| \ \| : V \to \mathbb{R}$ is the value of the function $d : V \times V \to \mathbb{R}$ given by*

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|.$$

Theorem 6.2.7. *The norm distance in Definition 6.2.6 satisfies for every $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z} \in V$ that*

(a) $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$, *and* $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ *iff* $\boldsymbol{x} = \boldsymbol{y}$, *(Positive definiteness);*
(b) $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$, *(Symmetry);*
(c) $d(\boldsymbol{x}, \boldsymbol{y}) \leqslant d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$, *(Triangle inequality).*

Proof of Theorem 6.2.7: Properties (a) and (b) are straightforward from properties (a) and (b), and their proof are left as an exercise. We show how the triangle inequality for the distance comes from the triangle inequality for the norm. Indeed

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \|(\boldsymbol{x} - \boldsymbol{z}) - (\boldsymbol{y} - \boldsymbol{z})\| \leqslant \|\boldsymbol{x} - \boldsymbol{z}\| + \|\boldsymbol{y} - \boldsymbol{z}\| = d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y}),$$

where we used the symmetry of the distance function on the last term above. This establishes the Theorem. □

The presence of an inner product, and hence a norm and a distance, on a vector space permits to introduce the notion of convergence of an infinite sequence of vectors. We say that the sequence $\{\boldsymbol{x}_n\}_{n=0}^{\infty} \subset V$ *converges* to $\boldsymbol{x} \in V$ iff

$$\lim_{n\to\infty} d(\boldsymbol{x}_n, \boldsymbol{x}) = 0.$$

Some of the most important concepts related to convergence are closeness of a subspace, completeness of the vector space, and the continuity of linear operators and linear transformations. In the case of finite dimensional vector spaces the situation is straightforward. All subspaces are closed, all inner product spaces are complete and all linear operators and linear transformations are continuous. However, in the case of infinite dimensional vector spaces, things are not so simple.

6.2.4. **Exercises.**

6.2.1.- Determine which of the following functions $\langle\,,\rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ defines an inner product on $\mathbb{R}^3$. Justify your answers.

(a) $\langle \mathsf{x}, \mathsf{y}\rangle = x_1y_1 + x_3y_3$;
(b) $\langle \mathsf{x}, \mathsf{y}\rangle = x_1y_1 - x_2y_2 + x_3y_3$;
(c) $\langle \mathsf{x}, \mathsf{y}\rangle = 2x_1y_1 + x_2y_2 + 4x_3y_3$;
(d) $\langle \mathsf{x}, \mathsf{y}\rangle = x_1^2\,y_1^2 + x_2^2\,y_2^2 + x_3^2\,y_3^2$.

We used the standard notation

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

6.2.2.- Prove that an inner product function $\langle\,,\rangle : V \times V \to \mathbb{F}$ satisfies the following properties:

(a) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$ for all $\boldsymbol{x} \in V$, then $\boldsymbol{y} = \boldsymbol{0}$.
(b) $\langle a\boldsymbol{x}, \boldsymbol{y}\rangle = \overline{a}\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}$, $\boldsymbol{y} \in V$.

6.2.3.- Given a matrix $\mathsf{M} \in \mathbb{R}^{2,2}$ introduce the function $\langle\,,\rangle_M : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

$$\langle \mathsf{y}, \mathsf{x}\rangle_M = \mathsf{y}^T\mathsf{M}\,\mathsf{x}.$$

For each of the matrices M below determine whether $\langle\,,\rangle_M$ defines an inner product or not. Justify your answers.

(a) $\mathsf{M} = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}$;
(b) $\mathsf{M} = \begin{bmatrix} 4 & -3 \\ 3 & 9 \end{bmatrix}$;
(c) $\mathsf{M} = \begin{bmatrix} 4 & 1 \\ 0 & 9 \end{bmatrix}$.

6.2.4.- Fix any $\mathsf{A} \in \mathbb{R}^{n,n}$ with $N(\mathsf{A}) = \{\mathsf{0}\}$ and introduce $\mathsf{M} = \mathsf{A}^T\mathsf{A}$. Prove that $\langle\,,\rangle_M : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, given by

$$\langle \mathsf{y}, \mathsf{x}\rangle = \mathsf{y}^T\mathsf{M}\,\mathsf{x}.$$

is an inner product in $\mathbb{R}^n$.

6.2.5.- Find $k \in \mathbb{R}$ such that the matrices A, $\mathsf{B} \in \mathbb{R}^{2,2}$ are perpendicular in the Frobenius inner product,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \; \mathsf{B} = \begin{bmatrix} 1 & 2 \\ k & 1 \end{bmatrix}.$$

6.2.6.- Evaluate the Frobenius norm for the matrices

$$\mathsf{A} = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

6.2.7.- Prove that $\|\mathsf{A}\|_F = \|\mathsf{A}^*\|_F$ for all $\mathsf{A} \in \mathbb{F}^{m,n}$.

6.2.8.- Consider the vector space $\mathbb{P}_2([0,1])$ with inner product

$$\langle \boldsymbol{p}, \boldsymbol{q}\rangle = \int_0^1 \boldsymbol{p}(x)\,\boldsymbol{q}(x)\,dx.$$

Find a unit vector parallel to

$$\boldsymbol{p}(x) = 3 - 5x^2.$$

6.3. Orthogonal vectors

6.3.1. **Definition and examples.** In the previous Section we introduced the notion of inner product in a vector space. This structure provides a notion of vector norm and distance between vectors. In this Section we explore another concept provided by an inner product; the notion of perpendicular vectors and the notion of angle between vectors in a real vector space. We start defining perpendicular vectors on any inner product space.

Definition 6.3.1. *Two vectors* $\boldsymbol{x}$, $\boldsymbol{y}$ *in and inner product space* $\big(V, \langle\,,\rangle\big)$ *are* ***orthogonal*** *or* ***perpendicular****, denoted as* $\boldsymbol{x} \perp \boldsymbol{y}$*, iff holds* $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$.

The Pythagoras Theorem holds on any inner product space.

Theorem 6.3.2. *Let* $\big(V, \langle\,,\rangle\big)$ *be an inner product space over the field* $\mathbb{F}$.

(a) If $\mathbb{F} = \mathbb{R}$, *then* $\boldsymbol{x} \perp \boldsymbol{y} \quad \Leftrightarrow \quad \|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$;
(b) If $\mathbb{F} = \mathbb{C}$, *then* $\boldsymbol{x} \perp \boldsymbol{y} \quad \Rightarrow \quad \|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

Proof of Theorem 6.3.2: Both statements derive from the following equation:

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \langle (\boldsymbol{x} - \boldsymbol{y}), (\boldsymbol{x} - \boldsymbol{y})\rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle - \langle \boldsymbol{y}, \boldsymbol{x}\rangle \\ &= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\,\mathsf{Re}\big(\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big). \end{aligned} \tag{6.5}$$

In the case $\mathbb{F} = \mathbb{R}$ holds $\mathsf{Re}\big(\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big) = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$, so Part (a) follows. If $\mathbb{F} = \mathbb{C}$, then $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ implies $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, so Part (b) follows. (Notice that the converse statement is not true in the case $\mathbb{F} = \mathbb{C}$, since Eq. (6.5) together with the hypothesis $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ do not fix $\mathsf{Im}\big(\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big)$.) This establishes the Theorem. □

In the case of real vector space the Cauchy-Schwarz inequality stated in Theorem 6.2.2 allows us to define the angle between vectors.

Definition 6.3.3. *The* ***angle*** *between two vectors* $\boldsymbol{x}$, $\boldsymbol{y}$ *in a real vector inner product space* $\big(V, \langle\,,\rangle\big)$ *is the number* $\theta \in [0, \pi]$ *solution of*

$$\cos(\theta) = \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}.$$

Example 6.3.1: Consider the inner product space $\big(\mathbb{R}^{2,2}, \langle\,,\rangle_F\big)$ and show that the following matrices are orthogonal,

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} -5 & 2 \\ 5 & 1 \end{bmatrix}.$$

Solution: Since we need to compute the Frobenius inner product $\langle \mathsf{A}, \mathsf{B}\rangle_F$, we first compute the matrix

$$\mathsf{A}^T\mathsf{B} = \begin{bmatrix} 1 & -1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -5 & 2 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} -10 & 1 \\ 5 & 10 \end{bmatrix}.$$

Therefore $\langle \mathsf{A}, \mathsf{B}\rangle_F = \mathrm{tr}\,\big(\mathsf{A}^T\mathsf{B}\big) = 0$, so we conclude that $\mathsf{A} \perp \mathsf{B}$. ◁

Example 6.3.2: Consider the vector space $V = C^\infty\big([-\ell, \ell], \mathbb{R}\big)$ with the inner product

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx.$$

Consider the functions $\boldsymbol{u}_n(x) = \cos\big(\frac{n\pi x}{\ell}\big)$ and $\boldsymbol{v}_m(x) = \sin\big(\frac{m\pi x}{\ell}\big)$, where n, m are integers.

(a) Show that $\boldsymbol{u}_n \perp \boldsymbol{v}_m$ for all n, m.
(b) Show that $\boldsymbol{u}_n \perp \boldsymbol{u}_m$ for all $n \neq m$.

(c) Show that $\boldsymbol{v}_n \perp \boldsymbol{v}_m$ for all $n \neq m$.

SOLUTION: Recall the identities

$$\sin(\theta)\cos(\phi) = \frac{1}{2}\big[\sin(\theta-\phi) + \sin(\theta+\phi)\big], \tag{6.6}$$

$$\cos(\theta)\cos(\phi) = \frac{1}{2}\big[\cos(\theta-\phi) + \cos(\theta+\phi)\big], \tag{6.7}$$

$$\sin(\theta)\sin(\phi) = \frac{1}{2}\big[\cos(\theta-\phi) - \cos(\theta+\phi)\big]. \tag{6.8}$$

Part (a): Using identity in Eq. (6.6) is simple to show that

$$\begin{aligned}\langle \boldsymbol{u}_n, \boldsymbol{v}_m\rangle &= \int_{-\ell}^{\ell} \cos\Big(\frac{n\pi x}{\ell}\Big)\,\sin\Big(\frac{m\pi x}{\ell}\Big)\,dx \\ &= \frac{1}{2}\int_{-\ell}^{\ell}\Big[\sin\Big(\frac{(n-m)\pi x}{\ell}\Big) + \sin\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big].\end{aligned} \tag{6.9}$$

First, assume that both $n-m$ and $n+m$ are non-zero,

$$\langle \boldsymbol{u}_n, \boldsymbol{v}_m\rangle = -\frac{1}{2}\Big[\frac{\ell}{(n-m)\pi}\,\cos\Big(\frac{(n-m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell} + \frac{\ell}{(n+m)\pi}\,\cos\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell}\Big]. \tag{6.10}$$

Since $\cos((n\pm m)\pi) = \cos(-(n\pm m)\pi)$, we conclude that both terms above vanish.

Second, in the case that $n-m=0$ the first term in Eq. (6.9) vanishes identically and we need to compute the term with $(n+m)$, which also vanishes by the second term in Eq. (6.10). Analogously, in the case of $(n+m)=0$ the second term in Eq. (6.9) vanishes identically and we need to compute the term with $(n-m)$ which also vanishes by the first term in Eq. (6.10). Therefore, $\langle \boldsymbol{u}_n, \boldsymbol{v}_m\rangle = 0$ for all n, m integers, and so $\boldsymbol{u}_n \perp \boldsymbol{v}_m$ in this case.

Part (b): Using identity in Eq. (6.7) is simple to show that

$$\begin{aligned}\langle \boldsymbol{u}_n, \boldsymbol{u}_m\rangle &= \int_{-\ell}^{\ell} \cos\Big(\frac{n\pi x}{\ell}\Big)\,\cos\Big(\frac{m\pi x}{\ell}\Big)\,dx \\ &= \frac{1}{2}\int_{-\ell}^{\ell}\Big[\cos\Big(\frac{(n-m)\pi x}{\ell}\Big) + \cos\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big].\end{aligned} \tag{6.11}$$

We know that $n-m$ is non-zero. Now, assume that $n+m$ is non-zero, then

$$\begin{aligned}\langle \boldsymbol{u}_n, \boldsymbol{u}_m\rangle &= \frac{1}{2}\Big[\frac{\ell}{(n-m)\pi}\,\sin\Big(\frac{(n-m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell} + \frac{\ell}{(n+m)\pi}\,\sin\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell}\Big] \\ &= \frac{\ell}{(n-m)\pi}\,\sin((n-m)\pi) + \frac{\ell}{(n+m)\pi}\,\sin((n+m)\pi).\end{aligned} \tag{6.12}$$

Since $\sin((n\pm m)\pi) = 0$ for $(n\pm m)$ integer, we conclude that both terms above vanish.

In the case of $(n+m)=0$ the second term in Eq. (6.11) vanishes identically and we need to compute the term with $(n-m)$ which also vanishes by the first term in Eq. (6.12). Therefore, $\langle \boldsymbol{u}_n, \boldsymbol{u}_m\rangle = 0$ for all $n \neq m$ integers, and so $\boldsymbol{u}_n \perp \boldsymbol{u}_m$ in this case.

Part (c): Using identity in Eq. (6.8) is simple to show that

$$\begin{aligned}\langle \boldsymbol{v}_n, \boldsymbol{v}_m\rangle &= \int_{-\ell}^{\ell} \sin\Big(\frac{n\pi x}{\ell}\Big)\,\sin\Big(\frac{m\pi x}{\ell}\Big)\,dx \\ &= \frac{1}{2}\int_{-\ell}^{\ell}\Big[\cos\Big(\frac{(n-m)\pi x}{\ell}\Big) - \cos\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big].\end{aligned} \tag{6.13}$$

Since the only difference between Eq. (6.13) and (6.11) is the sign of the second term, repeating the argument done in case (b) we conclude that $\langle \boldsymbol{v}_n, \boldsymbol{v}_m \rangle = 0$ for all $n \neq m$ integers, and so $\boldsymbol{v}_n \perp \boldsymbol{v}_m$ in this case. ◁

6.3.2. **Orthonormal basis.** We saw that an important property of a basis is that every vector in a vector space can be decomposed in a unique way in terms of the basis vectors. This decomposition is particularly simple to find in an inner product space when the basis is an orthonormal basis. Before we introduce such basis we define an orthonormal set.

Definition 6.3.4. *The set $U = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\}$, $p \geqslant 1$, in an inner product space $\big(V, \langle\,,\rangle\big)$ is called an* ***orthonormal set*** *iff for all $i, j = 1, \cdots, p$ holds*

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j. \end{cases}$$

The set U is called an ***orthogonal set*** *iff holds that $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0$ if $i \neq j$ and $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle \neq 0$.*

EXAMPLE 6.3.3: Consider the vector space $V = C^\infty\big([-\ell, \ell], \mathbb{R}\big)$ with the inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx.$$

Show that the set

$$U = \Big\{ \boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}},\ \boldsymbol{u}_n(x) = \frac{1}{\sqrt{\ell}} \cos\big(\frac{n\pi x}{\ell}\big),\ \boldsymbol{v}_m(x) = \frac{1}{\sqrt{\ell}} \sin\big(\frac{m\pi x}{\ell}\big) \Big\}_{n=1}^{\infty}$$

is an orthonormal set.

SOLUTION: We have shown in Example 6.3.2 that U is an orthogonal set. We only need to compute the norm of the vectors $\boldsymbol{u}_0$, $\boldsymbol{u}_n$ and $\boldsymbol{v}_n$, for $n = 1, 2, \cdots$. The norm of the first vector is simple to compute,

$$\langle \boldsymbol{u}_0, \boldsymbol{u}_0 \rangle = \int_{-\ell}^{\ell} \frac{1}{2\ell}\, dx = 1.$$

The norm of the cosine functions is computed as follows,

$$\begin{aligned} \langle \boldsymbol{u}_n, \boldsymbol{u}_n \rangle &= \frac{1}{\ell} \int_{-\ell}^{\ell} \cos^2\big(\frac{n\pi x}{\ell}\big)\, dx \\ &= \frac{1}{2\ell} \int_{-\ell}^{\ell} \Big[1 + \cos\big(\frac{2n\pi x}{\ell}\big)\Big]\, dx \\ &= 1 + \frac{\ell}{2n\pi} \Big[\sin\big(\frac{2n\pi x}{\ell}\big)\Big|_{-\ell}^{\ell}\Big] \quad \Rightarrow \quad \langle \boldsymbol{u}_n, \boldsymbol{u}_n \rangle = 1. \end{aligned}$$

A similar calculation for the sine functions gives the result

$$\begin{aligned} \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle &= \frac{1}{\ell} \int_{-\ell}^{\ell} \sin^2\big(\frac{n\pi x}{\ell}\big)\, dx \\ &= \frac{1}{2\ell} \int_{-\ell}^{\ell} \Big[1 - \cos\big(\frac{2n\pi x}{\ell}\big)\Big]\, dx \\ &= 1 - \frac{\ell}{2n\pi} \Big[\sin\big(\frac{2n\pi x}{\ell}\big)\Big|_{-\ell}^{\ell}\Big] \quad \Rightarrow \quad \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle = 1. \end{aligned}$$

Therefore, U is an orthonormal set. ◁

A straightforward result is the following:

Theorem 6.3.5. *An orthogonal set in an inner product space is linearly independent.*

Proof of Theorem 6.3.5: Let $U = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\}$, $p \geqslant 1$, be an orthogonal set. The zero vector is not included since $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle \neq 0$ for all $i = 1, \cdots, p$. Let $c_1, \cdots, c_p \in \mathbb{F}$ be scalars such that

$$c_1 \boldsymbol{u}_1 + \cdots + c_p \boldsymbol{u}_p = \boldsymbol{0}.$$

Then, for any $\boldsymbol{u}_i \in U$ holds

$$0 = \langle u_i, (c_1 \boldsymbol{u}_1 + \cdots + c_p \boldsymbol{u}_p \rangle = c_1 \langle \boldsymbol{u}_i, \boldsymbol{u}_1 \rangle + \cdots + c_p \langle \boldsymbol{u}_i, \boldsymbol{u}_p \rangle.$$

Since the set U is orthogonal, $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0$ for $i \neq j$ and $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle \neq 0$, so we conclude that

$$c_i \langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle = 0 \quad \Rightarrow \quad c_i = 0, \qquad i = 1, \cdots, p.$$

Therefore, U is a linearly independent set. This establishes the Thorem. □

Definition 6.3.6. *A basis $\mathcal{U}$ of an inner product space is called an **orthonormal basis** (**orthogonal basis**) iff the basis $\mathcal{U}$ is an orthonormal (orthogonal) set.*

EXAMPLE 6.3.4: Consider the inner product space $\big(\mathbb{R}^2, \cdot\big)$. Determine whether the following bases are orthonormal, orthogonal or neither:

$$\mathcal{S} = \left\{ \mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \quad \mathcal{U} = \left\{ \mathsf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}, \quad \mathcal{V} = \left\{ \mathsf{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \mathsf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}.$$

SOLUTION: The basis $\mathcal{S}$ is orthonormal, since $\mathsf{e}_1 \cdot \mathsf{e}_2 = 0$ and $\mathsf{e}_1 \cdot \mathsf{e}_1 = \mathsf{e}_2 \cdot \mathsf{e}_2 = 1$. The basis $\mathcal{U}$ is orthogonal since $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 0$, but it is not orthonormal. Finally, the basis $\mathcal{V}$ is neither orthonormal nor orthogonal. ◁

Theorem 6.3.7. *Given the set $U = \{\mathsf{u}_1, \cdots, \mathsf{u}_p\}$, $p \geqslant 1$, in the inner product space $\big(\mathbb{F}^n, \cdot\big)$, introduce the matrix $\mathsf{U} = [\mathsf{u}_1, \cdots, \mathsf{u}_p]$. Then the following statements hold:*

(a) U is an orthonormal set iff matrix U satisfies $\mathsf{U}^*\mathsf{U} = \mathsf{I}_p$.

(b) U is an orthonormal basis of $\mathbb{F}^n$ iff matrix U satisfies $\mathsf{U}^{-1} = \mathsf{U}^*$.

Matrices satisfying the property mentioned in part (b) of Theorem 6.3.7 appear frequently in Quantum Mechanics, so they are given a name.

Definition 6.3.8. *A matrix $\mathsf{U} \in \mathbb{F}^{n,n}$ is called **unitary** iff holds $\mathsf{U}^{-1} = \mathsf{U}^*$.*

These matrices are called unitary because they do not change the norm of vectors in $\mathbb{F}^n$ equipped with the dot product. Indeed, given any $\mathsf{x} \in \mathbb{F}^n$ with norm $\|\mathsf{x}\|$, the vector $\mathsf{U}\mathsf{x} \in \mathbb{F}^n$ has the same norm, since

$$\|\mathsf{U}\mathsf{x}\|^2 = \big(\mathsf{U}\mathsf{x}\big)^* \big(\mathsf{U}\mathsf{x}\big) = \mathsf{x}^*\mathsf{U}^*\mathsf{U}\mathsf{x} = \mathsf{x}^*\mathsf{U}^{-1}\mathsf{U}\mathsf{x} = \mathsf{x}^*\mathsf{x} = \|\mathsf{x}\|^2.$$

Proof of Theorem 6.3.7:

Part (a): This is proved by a straightforward computation,

$$\mathsf{U}^*\mathsf{U} = \begin{bmatrix} \mathsf{u}_1^* \\ \vdots \\ \mathsf{u}_p^* \end{bmatrix} \big[\mathsf{u}_1, \cdots, \mathsf{u}_p\big] = \begin{bmatrix} \mathsf{u}_1^*\mathsf{u}_1 & \cdots & \mathsf{u}_1^*\mathsf{u}_p \\ \vdots & & \vdots \\ \mathsf{u}_p^*\mathsf{u}_1 & \cdots & \mathsf{u}_p^*\mathsf{u}_p \end{bmatrix} = \begin{bmatrix} \mathsf{u}_1 \cdot \mathsf{u}_1 & \cdots & \mathsf{u}_1 \cdot \mathsf{u}_p \\ \vdots & & \vdots \\ \mathsf{u}_p \cdot \mathsf{u}_1 & \cdots & \mathsf{u}_p \cdot \mathsf{u}_p \end{bmatrix} = \mathsf{I}_p.$$

Part (b): It follows from part (a): If U is a basis of $\mathbb{F}^n$, then $p = n$; since U is an orthonormal set, part (a) implies $\mathsf{U}^*\mathsf{U} = \mathsf{I}_n$. Since U is an $n \times n$ matrix, it follows that $\mathsf{U}^* = \mathsf{U}^{-1}$. This establishes the Theorem. □

EXAMPLE 6.3.5: Consider $\mathsf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, $\mathsf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$, in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

(a) Show that $\mathsf{v}_1 \perp \mathsf{v}_2$;

(b) Find $\mathsf{x} \in \mathbb{R}^3$ such that $\mathsf{x} \perp \mathsf{v}_1$ and $\mathsf{x} \perp \mathsf{v}_2$.
(c) Rescale the elements of $\{\mathsf{v}_1, \mathsf{v}_2, \mathsf{x}\}$ so that the new set is an orthonormal set.

SOLUTION:

Part (a):

$$\begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = 2 + 0 - 2 = 0 \quad \Rightarrow \quad \mathsf{v}_1 \perp \mathsf{v}_2.$$

Part (b): We need to find $\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ such that

$$\mathsf{v}_1 \cdot \mathsf{x} = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0, \qquad \mathsf{v}_2 \cdot \mathsf{x} = \begin{bmatrix} 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0,$$

The equations above can be written in matrix notation as

$$\mathsf{A}^T \mathsf{x} = 0, \quad \text{where} \quad \mathsf{A} = \begin{bmatrix} \mathsf{v}_1, \mathsf{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}.$$

Gauss elimination operation on A^T imply

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & 5/2 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = \dfrac{1}{2}\, x_3, \\ x_2 = -\dfrac{5}{2}\, x_3, \\ x_3 \text{ free.} \end{cases}$$

There is a solution for any choice of $x_3 \neq 0$, so we choose $x_3 = 2$, that is, $\mathsf{x} = \begin{bmatrix} 1 \\ -5 \\ 2 \end{bmatrix}$.

Part (c): The vectors v_1, v_2 and x are mutually orthogonal. Their norms are:

$$\|\mathsf{v}_1\| = \sqrt{6}, \qquad \|\mathsf{v}_2\| = \sqrt{5}, \qquad \|\mathsf{x}\| = \sqrt{30}.$$

Therefore, the orthonormal set is

$$\left\{ \mathsf{u}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \; \mathsf{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \; \mathsf{u}_3 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ -5 \\ 2 \end{bmatrix} \right\}.$$

Finally, notice that the inverse of the matrix

$$\mathsf{U} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & -\frac{5}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \end{bmatrix} \quad \text{is} \quad \mathsf{U}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{30}} & -\frac{5}{\sqrt{30}} & \frac{2}{\sqrt{30}} \end{bmatrix} = \mathsf{U}^T.$$

◁

6.3.3. **Vector components.** Given a basis in a finite dimensional vector space, we know that every vector in the vector space can be decomposed in a unique way as a linear combination of the basis vectors. What we do not know is a general formula to compute the vector components in such a basis. However, in the particular case that the vector space admits an inner product and the basis is an orthonormal basis, we have such formula for the vector components. The formula is very simple and is the main result in the following statement.

Theorem 6.3.9. *If $\big(V, \langle\,,\rangle\big)$ is an n-dimensional inner product space with an orthonormal basis $\mathcal{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$, then every vector $\boldsymbol{x} \in V$ can be decomposed as*

$$\boldsymbol{x} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle\, \boldsymbol{u}_n. \tag{6.14}$$

Proof of Theorem 6.3.9: Since $\mathcal{U}$ is a basis, we know that for all $\boldsymbol{x} \in V$ there exist scalars $c_1, \cdots, c_n$ such that

$$\boldsymbol{x} = c_1\, \boldsymbol{u}_1 + \cdots + c_n\, \boldsymbol{u}_n.$$

Therefore, the inner product $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$ for any $i = 1, \cdots, n$ is given by

$$\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle = c_1 \langle \boldsymbol{u}_i, \boldsymbol{u}_1\rangle + \cdots + c_n \langle \boldsymbol{u}_i, \boldsymbol{u}_n\rangle.$$

Since $\mathcal{U}$ is an orthonormal set, $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle = c_i$. This establishes the Theorem. □

The result in Theorem 6.3.9 provides a remarkable simple formula for vector components in an orthonormal basis. We will get back to this subject in some depth in Section 7.1. In that Section we will name the coefficients $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$ in Eq. (6.14) as *Fourier coefficients* of the vector $\boldsymbol{x}$ in the orthonormal set $\mathcal{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$. When the set $\mathcal{U}$ is an orthonormal ordered basis of a vector space V, the coordinate map $[\ ]_u : V \to \mathbb{F}^n$ is expressed in terms of the Fourier coefficients as follows,

$$[\boldsymbol{x}]_u = \begin{bmatrix} \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle \\ \vdots \\ \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle \end{bmatrix}.$$

So, the coordinate map has a simple expression when it is defined by an orthonormal basis.

EXAMPLE 6.3.6: Consider the inner product space $\big(\mathbb{R}^3, \cdot\big)$ with the standard ordered basis $\mathcal{S}$, and find the vector components of $\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ in the orthonormal ordered basis

$$\mathcal{U} = \left(\mathsf{u}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix},\ \mathsf{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix},\ \mathsf{u}_3 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ -5 \\ 2 \end{bmatrix}\right).$$

SOLUTION: The vector components of x in the orthonormal basis $\mathcal{U}$ are given by

$$\mathsf{x}_u = \begin{bmatrix} \langle \mathsf{u}_1, \mathsf{x}\rangle \\ \langle \mathsf{u}_2, \mathsf{x}\rangle \\ \langle \mathsf{u}_3, \mathsf{x}\rangle \end{bmatrix} \quad \Rightarrow \quad \mathsf{u} = \begin{bmatrix} \frac{9}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} \\ -\frac{3}{\sqrt{30}} \end{bmatrix}.$$

REMARK: We have done a change of basis, from the standard basis $\mathcal{S}$ to the $\mathcal{U}$ basis. In fact, we can express the calculation above as follows

$$\mathsf{x}_u = \mathsf{P}^{-1}\mathsf{x}_s, \quad \text{where} \quad \mathsf{P} = \mathsf{I}_{us} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & -\frac{5}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \end{bmatrix} = \mathsf{U}.$$

Since $\mathcal{U}$ is an orthonormal basis, $\mathsf{U}^{-1} = \mathsf{U}^T$, so we conclude that

$$\mathsf{x}_u = \mathsf{U}^T\mathsf{x}_s = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{30}} & -\frac{5}{\sqrt{30}} & \frac{2}{\sqrt{30}} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{9}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} \\ -\frac{3}{\sqrt{30}} \end{bmatrix}.$$

◁

6.3.4. **Exercises.**

6.3.1.- Prove that the following form of the Pythagoras Theorem holds on complex vector spaces: Two vectors $\boldsymbol{x}$, $\boldsymbol{y}$ in an inner product space $\big(V, \langle\,,\rangle\big)$ over $\mathbb{C}$ are orthogonal iff for all a, $b \in \mathbb{C}$ holds

$$\|a\boldsymbol{x} + b\boldsymbol{y}\|^2 = \|a\boldsymbol{x}\|^2 + \|b\boldsymbol{y}\|^2.$$

6.3.2.- Consider the vector space $\mathbb{R}^3$ with the dot product. Find all vectors $\mathsf{x} \in \mathbb{R}^3$ which are orthogonal to the vector

$$\mathsf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

6.3.3.- Consider the vector space $\mathbb{R}^3$ with the dot product. Find all vectors $\mathsf{x} \in \mathbb{R}^3$ which are orthogonal to the vectors

$$\mathsf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

6.3.4.- Let $\mathbb{P}_3([-1,1])$ be the space of polynomials up to degree three defined on the interval $[-1,1] \subset \mathbb{R}$ with the inner product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} \boldsymbol{p}(x)\boldsymbol{q}(x)\, dx.$$

Show that the set $(\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)$ is an orthogonal basis of $\mathbb{P}_3$, where

$$\begin{aligned} \boldsymbol{p}_0(x) &= 1, \\ \boldsymbol{p}_1(x) &= x, \\ \boldsymbol{p}_2(x) &= \frac{1}{2}(3x^2 - 1), \\ \boldsymbol{p}_3(x) &= \frac{1}{2}(5x^3 - 3x). \end{aligned}$$

(These polynomials are the first four of the Legendre polynomials.)

6.3.5.- Consider the vector space $\mathbb{R}^3$ with the dot product.

(a) Show that the following ordered basis $\mathcal{U}$ is orthonormal,

$$\Big(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}\Big).$$

(b) Use part (a) to find the components in the ordered basis $\mathcal{U}$ of the vector

$$\mathsf{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

6.3.6.- Consider the vector space $\mathbb{R}^{2,2}$ with the Frobenius inner product.

(a) Show that the ordered basis given by $\mathcal{U} = (\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3, \mathsf{E}_4)$ is orthonormal, where

$$\mathsf{E}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \mathsf{E}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\mathsf{E}_3 = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \mathsf{E}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

(b) Use part (a) to find the components in the ordered basis $\mathcal{U}$ of the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

6.3.7.- Consider the inner product space $\big(\mathbb{R}^{2,2}, \langle\,,\rangle_F\big)$, and find the cosine of the angle between the matrices

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 2 & -2 \\ 2 & 0 \end{bmatrix}.$$

6.3.8.- Find the third column in matrix U below such that $\mathsf{U}^T = \mathsf{U}^{-1}$, where

$$\mathsf{U} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{14} & U_{13} \\ 1/\sqrt{3} & 2/\sqrt{14} & U_{23} \\ 1/\sqrt{3} & -3/\sqrt{14} & U_{33} \end{bmatrix}.$$

6.4. Orthogonal projections

6.4.1. **Orthogonal projection onto subspaces.** Given any subspace of an inner product space, every vector in the vector space can be decomposed as a sum of two vectors; a vector in the subspace and a vector perpendicular to the subspace. The picture one often has in mind is a plane in $\mathbb{R}^3$ equiped with the dot product, as in Fig. 43. Any vector in $\mathbb{R}^3$ can be decomposed as a vector on the plane plus a vector perpendicular to the plane. In this Section we provide expressions to compute this type of decompositions. We start splitting a vector in orthogonal components with respect to a one dimensional subspace. The study of this simple case describes the main ideas and the main notation used in orthogonal decompositions. Later on we present the decomposition of a vector onto an n-dimensional subspace.

Theorem 6.4.1. *Fix a vector $\boldsymbol{u} \neq \mathbf{0}$ in an inner product space $\big(V, \langle\,,\rangle\big)$. Given any vector $\boldsymbol{x} \in V$ decompose it as $\boldsymbol{x} = \boldsymbol{x}_{//} + \boldsymbol{x}_{\perp}$ where $\boldsymbol{x}_{//} \in \text{Span}(\{\boldsymbol{u}\})$. Then, $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}$ iff holds*

$$\boldsymbol{x}_{//} = \frac{\langle \boldsymbol{u}, \boldsymbol{x}\rangle}{\|\boldsymbol{u}\|^2}\,\boldsymbol{u}. \tag{6.15}$$

Furthermore, in the case that $\boldsymbol{u}$ is a unit vector holds $\boldsymbol{x}_{//} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle\,\boldsymbol{u}$.

The main idea of this decomposition can be understood in the inner product space $\big(\mathbb{R}^2, \cdot\big)$ and it is sketched in Fig. 42. It is obvious that every vector $\boldsymbol{x}$ can be expressed in many different ways as the sum of two vectors. What is special of the decomposition in Eq. 6.15 is that $\boldsymbol{x}_{//}$ has the precise length such that $\boldsymbol{x}_{\perp}$ is orthogonal to $\boldsymbol{x}_{//}$ (see Fig. 42).

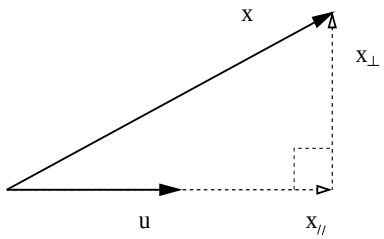

FIGURE 42. Orthogonal decomposition of the vector $\mathsf{x} \in \mathbb{R}^2$ onto the subspace spanned by vector u.

Proof of Theorem 6.4.1: Since $\boldsymbol{x}_{//} \in \text{Span}(\{\boldsymbol{u}\})$, there exists a scalar a such that $\boldsymbol{x}_{//} = a\boldsymbol{u}$. Therefore $\boldsymbol{u} \perp \boldsymbol{x}_{\perp}$ iff holds that $\langle \boldsymbol{u}, \boldsymbol{x}_{\perp}\rangle = 0$. A straightforward computation shows,

$$0 = \langle \boldsymbol{u}, \boldsymbol{x}_{\perp}\rangle = \langle \boldsymbol{u}, \boldsymbol{x}\rangle - \langle \boldsymbol{u}, \boldsymbol{x}_{//}\rangle = \langle \boldsymbol{u}, \boldsymbol{x}\rangle - a\,\langle \boldsymbol{u}, \boldsymbol{u}\rangle \quad \Leftrightarrow \quad a = \frac{\langle \boldsymbol{u}, \boldsymbol{x}\rangle}{\|\boldsymbol{u}\|^2}.$$

We conclude that the decomposition $\boldsymbol{x} = \boldsymbol{x}_{//} + \boldsymbol{x}_{\perp}$ satisfies

$$\boldsymbol{x}_{\perp} \perp \boldsymbol{u} \quad \Leftrightarrow \quad \boldsymbol{x}_{//} = \frac{\langle \boldsymbol{u}, \boldsymbol{x}\rangle}{\|\boldsymbol{u}\|^2}\,\boldsymbol{u}.$$

In the case that $\boldsymbol{u}$ is a unit vector holds $\|\boldsymbol{u}\| = 1$, so $\boldsymbol{x}_{//}$ is given by $\boldsymbol{x}_{//} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle\,\boldsymbol{u}$. This establishes the Theorem. □

EXAMPLE 6.4.1: Consider the inner product space $\big(\mathbb{R}^3, \cdot\big)$ and decompose the vector x in orthogonal components with respect to the vector u, where

$$\mathsf{x} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \qquad \mathsf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

SOLUTION: We first compute $\mathsf{x}_{\shortparallel} = \dfrac{\mathsf{u}\cdot\mathsf{x}}{\|\mathsf{u}\|^2}\,\mathsf{u}$. Since

$$\mathsf{u}\cdot\mathsf{x} = \begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}3\\2\\1\end{bmatrix} = 3+4+3 = 10, \quad \|\mathsf{u}\|^2 = \begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = 1+4+9 = 14,$$

we obtain $\mathsf{x}_{\shortparallel} = \dfrac{5}{7}\begin{bmatrix}1\\2\\3\end{bmatrix}$. We now compute $\mathsf{x}_{\perp}$ as follows,

$$\mathsf{x}_{\perp} = \mathsf{x} - \mathsf{x}_{\shortparallel} = \begin{bmatrix}3\\2\\1\end{bmatrix} - \frac{5}{7}\begin{bmatrix}1\\2\\3\end{bmatrix} = \frac{1}{7}\begin{bmatrix}21\\14\\7\end{bmatrix} - \frac{1}{7}\begin{bmatrix}5\\10\\15\end{bmatrix} = \frac{1}{7}\begin{bmatrix}16\\4\\-8\end{bmatrix} \quad \Rightarrow \quad \mathsf{x}_{\perp} = \frac{4}{7}\begin{bmatrix}4\\1\\-2\end{bmatrix}.$$

Therefore, x can be decomposed as

$$\mathsf{x} = \frac{5}{7}\begin{bmatrix}1\\2\\3\end{bmatrix} + \frac{4}{7}\begin{bmatrix}4\\1\\-2\end{bmatrix}.$$

REMARK: We can verify that this decomposition is orthogonal with respect to u, since

$$\mathsf{u}\cdot\mathsf{x}_{\perp} = \frac{4}{7}\begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}4\\1\\-2\end{bmatrix} = \frac{4}{7}(4+2-6) = 0.$$

◁

We now decompose a vector into orthogonal components with respect to a p-dimensional subspace with $p \geqslant 1$.

Theorem 6.4.2. *Fix an orthogonal set $\mathcal{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\}$, with $p \geqslant 1$, in an inner product space $\big(V, \langle\,,\rangle\big)$. Given any vector $\boldsymbol{x} \in V$, decompose it as $\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_{\perp}$, where $\boldsymbol{x}_{\shortparallel} \in \mathrm{Span}(\mathcal{U})$. Then, $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}_i$, for $i = 1, \cdots, p$ iff holds*

$$\boldsymbol{x}_{\shortparallel} = \frac{\langle \boldsymbol{u}_1, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_1\|^2}\,\boldsymbol{u}_1 + \cdots + \frac{\langle \boldsymbol{u}_p, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_p\|^2}\,\boldsymbol{u}_p. \tag{6.16}$$

Furthermore, in the case that $\mathcal{U}$ is an orthonormal basis holds

$$\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\,\boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_p, \boldsymbol{x}\rangle\,\boldsymbol{u}_p. \tag{6.17}$$

REMARK: The set $\mathcal{U}$ in Theorem 6.4.2 must be an orthogonal set. If the vectors $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p$ are not mutually orthogonal, then the vector $\boldsymbol{x}_{\shortparallel}$ computed in Eq. (6.17) is not the orthogonal projection of vector $\boldsymbol{x}$, that is, $(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) \not\perp \boldsymbol{u}_i$ for $i = 1, \cdots, p$. Therefore, before using Eq. (6.17) in a particular application one should verify that the set $\mathcal{U}$ one is working with is in fact an orthogonal set. A particular case of the orthogonal projection of a vector in the inner product space $\big(\mathbb{R}^3, \cdot\big)$ onto a plane is sketched in Fig. 43.

Proof of Theorem 6.4.2: Since $\boldsymbol{x}_{\shortparallel} \in \mathrm{Span}(\mathcal{U})$, there exist scalars a_i, for $i = 1, \cdots, p$ such that $\boldsymbol{x}_{\shortparallel} = a_1\boldsymbol{u}_1 + \cdots + a_p\boldsymbol{u}_p$. The vector $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}_i$ iff holds that $\langle \boldsymbol{u}_i, \boldsymbol{x}_{\perp}\rangle = 0$. A straightforward computation shows that, for $i = 1, \cdots, p$ holds

$$\begin{aligned} 0 = \langle \boldsymbol{u}_i, \boldsymbol{x}_{\perp}\rangle &= \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - \langle \boldsymbol{u}_i, \boldsymbol{x}_{\shortparallel}\rangle \\ &= \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - a_1\langle \boldsymbol{u}_i, \boldsymbol{u}_1\rangle - \cdots - a_p\langle \boldsymbol{u}_i, \boldsymbol{u}_p\rangle \\ &= \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - a_i\langle \boldsymbol{u}_i, \boldsymbol{u}_i\rangle \quad \Leftrightarrow \quad a_i = \frac{\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_i\|^2}. \end{aligned}$$

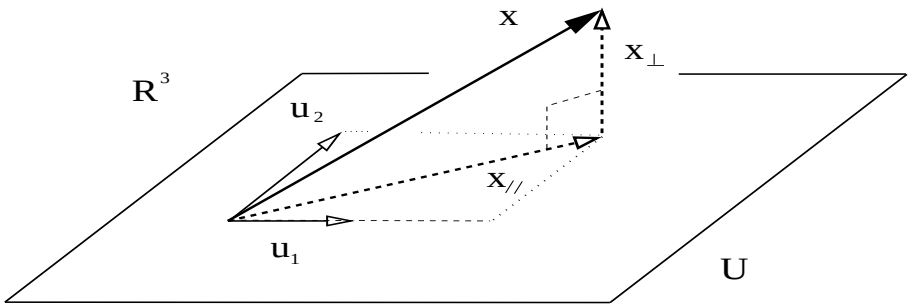

FIGURE 43. Orthogonal decomposition of the vector $\mathsf{x} \in \mathbb{R}^3$ onto the subspace U spanned by the vectors u_1 and u_2.

We conclude that the decomposition $\boldsymbol{x} = \boldsymbol{x}_{//} + \boldsymbol{x}_{\perp}$ satisfies $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}_i$ for $i = 1, \cdots, p$ iff holds

$$\boldsymbol{x}_{//} = \frac{\langle \boldsymbol{u}_1, \boldsymbol{x} \rangle}{\|\boldsymbol{u}_1\|^2}\, \boldsymbol{u}_1 + \cdots + \frac{\langle \boldsymbol{u}_p, \boldsymbol{x} \rangle}{\|\boldsymbol{u}_p\|^2}\, \boldsymbol{u}_p.$$

In the case that $\mathcal{U}$ is an orthonormal set holds $\|\boldsymbol{u}_i\| = 1$ for $i = 1, \cdots, p$, so $\boldsymbol{x}_{//}$ is given by Eq. (6.17). This establishes the Theorem. □

EXAMPLE 6.4.2: Consider the inner product space $\big(\mathbb{R}^3, \cdot\big)$ and decompose the vector x in orthogonal components with respect to the subspace U, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad U = \text{Span}\Big(\Big\{ \mathsf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \Big\}\Big).$$

SOLUTION: In order to use Eq. (6.16) we need an orthogonal basis of U. So, need need to verify whether u_1 is orthogonal to u_2. This is indeed the case, since

$$\mathsf{u}_1 \cdot \mathsf{u}_2 = \begin{bmatrix} 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = -4 + 5 - 1 = 0.$$

So now we use u_1 and u_2 to compute $\mathsf{x}_{//}$ using Eq. (6.17). We need the quantities

$$\mathsf{u}_1 \cdot \mathsf{x} = \begin{bmatrix} 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 9, \qquad \mathsf{u}_1 \cdot \mathsf{u}_1 = \begin{bmatrix} 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} = 30,$$

$$\mathsf{u}_2 \cdot \mathsf{x} = \begin{bmatrix} -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 3, \qquad \mathsf{u}_2 \cdot \mathsf{u}_2 = \begin{bmatrix} -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = 6.$$

Now is simple to compute $\mathsf{x}_{//}$, since

$$\mathsf{x}_{//} = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{3}{10} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 6 \\ 15 \\ -3 \end{bmatrix} + \frac{1}{10} \begin{bmatrix} -10 \\ 5 \\ 5 \end{bmatrix},$$

therefore,

$$\mathsf{x}_{//} = \frac{1}{10} \begin{bmatrix} -4 \\ 20 \\ 2 \end{bmatrix} \quad \Rightarrow \quad \mathsf{x}_{//} = \frac{1}{5} \begin{bmatrix} -2 \\ 10 \\ 1 \end{bmatrix}.$$

The vector $x_\perp$ is obtained as follows,

$$x_\perp = x - x_{/\!/} = \begin{bmatrix}1\\2\\3\end{bmatrix} - \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix} = \frac{1}{5}\begin{bmatrix}5\\10\\15\end{bmatrix} - \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix} = \frac{1}{5}\begin{bmatrix}7\\0\\14\end{bmatrix} \quad \Rightarrow \quad x_\perp = \frac{7}{5}\begin{bmatrix}1\\0\\2\end{bmatrix}.$$

We conclude that

$$x = \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix} + \frac{7}{5}\begin{bmatrix}1\\0\\2\end{bmatrix}.$$

REMARK: We can verify that $x_\perp \perp U$, since

$$u_1 \cdot x_\perp = \frac{7}{5}\begin{bmatrix}2 & 5 & -1\end{bmatrix}\begin{bmatrix}1\\0\\2\end{bmatrix} = \frac{7}{5}(2+0-2) = 0,$$

$$u_2 \cdot x_\perp = \frac{7}{5}\begin{bmatrix}-2 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\0\\2\end{bmatrix} = \frac{7}{5}(2+0-2) = 0.$$

◁

6.4.2. **Orthogonal complement.** We have seen that any vector in an inner product space can be decomposed as a sum of orthogonal vectors, that is, $\boldsymbol{x} = \boldsymbol{x}_{/\!/} + \boldsymbol{x}_\perp$ with $\langle \boldsymbol{x}_{/\!/}, \boldsymbol{x}_\perp \rangle = 0$. Inner product spaces can be decomposed in a somehow analogous way as a direct sum of two mutually orthogonal subspaces. In order to understand such decomposition, we need to introduce the notion of the orthogonal complement of a subspace. Then we will show that a finite dimensional inner product space is the direct sum of a subspace and it orthogonal complement. It is precisely this result that motivates the word "complement" in the name orthogonal complement of a subspace.

Definition 6.4.3. *The **orthogonal complement** of a subspace W in an inner product space $\big(V, \langle\,,\rangle\big)$, denoted as $W^\perp$, is the set $W^\perp = \{\boldsymbol{u} \in V \;:\; \langle \boldsymbol{u}, \boldsymbol{w}\rangle = 0 \quad \forall\ \boldsymbol{w} \in W\}$.*

EXAMPLE 6.4.3: In the inner product space $\big(\mathbb{R}^3, \cdot\big)$, the orthogonal complement to a line is a plane, and the orthogonal complement to a plane is a line, as it is shown in Fig. 44. ◁

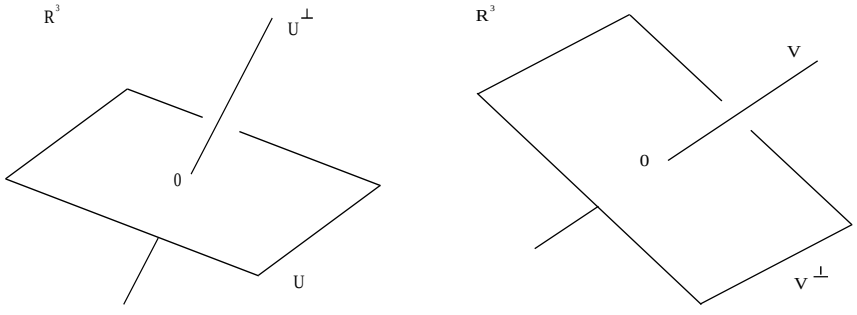

FIGURE 44. The orthogonal complement to the plane U is the line $U^\perp$, and the orthogonal complement to the line V is the plane $V^\perp$.

As the sketch in Fig. 44 suggests, the orthogonal complement of a subspace is a subspace.

Theorem 6.4.4. *The orthogonal complement $W^\perp$ of a subspace W in an inner product space is also a subspace.*

Proof of Theorem 6.4.4: Let $\boldsymbol{u}_1,\ \boldsymbol{u}_2 \in W^\perp$, that is, $\langle \boldsymbol{u}_i, \boldsymbol{w}\rangle = 0$ for all $\boldsymbol{w} \in W$ and $i = 1, 2$. Then, any linear combination $a\boldsymbol{u}_1 + b\boldsymbol{u}_2$ also belongs to $W^\perp$, since

$$\langle (a\boldsymbol{u}_1 + b\boldsymbol{u}_2), \boldsymbol{w}\rangle = \overline{a}\,\langle \boldsymbol{u}_1, \boldsymbol{w}\rangle + \overline{b}\,\langle \boldsymbol{u}_2, \boldsymbol{w}\rangle = 0 + 0 \qquad \forall \boldsymbol{w} \in W.$$

This establishes the Theorem. □

EXAMPLE 6.4.4: Find $W^\perp$ for the subspace $W = \text{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}\Big\}\Big)$ in $(\mathbb{R}^3, \cdot)$.

SOLUTION: We need to find the set of all $\mathsf{x} \in \mathbb{R}^3$ such that $\mathsf{x} \cdot \mathsf{w}_1 = 0$. That is,

$$\begin{bmatrix} -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \quad \Rightarrow \quad x_1 = 2x_2 + 3x_3.$$

The solution is

$$\mathsf{x} = \begin{bmatrix} 2x_2 + 3x_3 \\ x_2 \\ x_3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} x_3,$$

hence we obtain

$$W^\perp = \text{Span}\Big(\Big\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix},\ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}\Big\}\Big).$$

The orthogonal complement of a line is a plane, as sketched in the second picture in Fig. 44.
REMARK: We can verify that the result is correct, since

$$\begin{bmatrix} -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = -2 + 2 + 0 = 0, \qquad \begin{bmatrix} -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = -3 + 0 + 3 = 0.$$

◁

EXAMPLE 6.4.5: Find $W^\perp$ for the subspace $W = \text{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},\ \mathsf{w}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}\Big\}\Big)$ in $(\mathbb{R}^3, \cdot)$.

SOLUTION: We need to find the set of all $\mathsf{x} \in \mathbb{R}^3$ such that $\mathsf{x} \cdot \mathsf{w}_1 = 0$ and $\mathsf{x} \cdot \mathsf{w}_2 = 0$. That is,

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

We an use Gauss elimination to find the solution,

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = x_3, \\ x_2 = -2x_3, \\ x_3 \text{ free.} \end{cases} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} x_3,$$

hence we obtain

$$W^\perp = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\Big\}\Big).$$

So, the orthogonal complement of a pane is a line, as sketched in the first picture in Fig. 44.

REMARK: We can verify that the result is correct, since

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 1 - 4 + 3 = 0, \qquad \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 3 - 4 + 1 = 0.$$

◁

As we mentioned above, the reason for the word "complement" in the name of an orthogonal complement is that the vector space can be split into the sum of two subspaces with zero intersection. We summarize this below.

Theorem 6.4.5. *If W is a subspace in a finite dimensional inner product space V, then*

$$V = W \oplus W^{\perp}.$$

Proof of Theorem 6.4.5: We first show that $V = W + W^{\perp}$. In order to do that, first choose an orthonormal basis $\mathcal{W} = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_p\}$ for W, we here we have assumed $\dim W = p \leqslant n = \dim V$. Then, for every vector $\boldsymbol{x} \in V$ holds that it can be decomposed as

$$\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_{\perp}, \qquad \boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{w}_1, \boldsymbol{x} \rangle \, \boldsymbol{w}_1 + \cdots + \langle \boldsymbol{w}_p, \boldsymbol{x} \rangle \, \boldsymbol{w}_p, \qquad \boldsymbol{x}_{\perp} = \boldsymbol{x} - \boldsymbol{x}_{\shortparallel}.$$

Theorem 6.4.2 says that $\boldsymbol{x}_{\perp} \perp \boldsymbol{w}_i$ for $i = 1, \cdots, p$. This implies that $\boldsymbol{x}_{\perp} \in W^{\perp}$ and, since $\boldsymbol{x}$ is an arbitrary vector in V, we have established that $V = W + W^{\perp}$. We now show that $W \cap W^{\perp} = \{\boldsymbol{0}\}$. Indeed, if $\boldsymbol{u} \in W^{\perp}$, then $\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{w} \in W$. If $\boldsymbol{u} \in W$, then choosing $\boldsymbol{w} = \boldsymbol{u}$ in the equation above we get $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$, which implies $\boldsymbol{u} = 0$. Therefore, $W \cap W^{\perp} = \{\boldsymbol{0}\}$ and we conclude that $V = W \oplus W^{\perp}$. This establishes the Theorem. □

Since the orthogonal complement $W^{\perp}$ of a subspace W is itself a subspace, we can compute $(W^{\perp})^{\perp}$. The following statement says that the result is the original subspace W.

Theorem 6.4.6. *If W is a subspace in an finite dimensional inner product space, then*

$$\big(W^{\perp}\big)^{\perp} = W.$$

Proof of Theorem 6.4.6: First notice that $W \subset \big(W^{\perp}\big)^{\perp}$. Indeed, given any fixed vector $\boldsymbol{w} \in W$, the definition of $W^{\perp}$ implies that every vector $\boldsymbol{u} \in W^{\perp}$ satisfies $\langle \boldsymbol{w}, \boldsymbol{u} \rangle = 0$. This condition says that $\boldsymbol{w} \in \big(W^{\perp}\big)^{\perp}$, hence we conclude $W \subset \big(W^{\perp}\big)^{\perp}$. Second, Theorem 6.4.2 says that

$$V = W^{\perp} \oplus \big(W^{\perp}\big)^{\perp}.$$

In particular, this decomposition implies that $\dim V = \dim W^{\perp} + \dim\big(W^{\perp}\big)^{\perp}$. Again Theorem 6.4.2 also says that

$$V = W \oplus W^{\perp},$$

which in particular implies that $\dim V = \dim W + \dim W^{\perp}$. These last two results put together imply $\dim W = \dim\big(W^{\perp}\big)^{\perp}$. It is from this last result together with our previous result, $W \subset \big(W^{\perp}\big)^{\perp}$, that we obtain $W = \big(W^{\perp}\big)^{\perp}$. This establishes the Theorem. □

6.4.3. **Exercises.**

6.4.1.- Consider the inner product space $\big(\mathbb{R}^3, \cdot\big)$ and use Theorem 6.4.1 to find the orthogonal decomposition of vector x along vector u, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

6.4.2.- Consider the subspace W given by

$$\text{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathsf{w}_2 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}\Big\}\Big)$$

in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

(a) Find an orthogonal decomposition of the vector w_2 with respect to the vector w_1. Using this decomposition, find an orthogonal basis for the space W.

(b) Find the decomposition of the vector x below in orthogonal components with respect to the subspace W, where

$$\mathsf{x} = \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}.$$

6.4.3.- Consider the subspace W given by

$$\text{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}, \ \mathsf{w}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \Big\}\Big)$$

in the inner product space $\big(\mathbb{R}^3, \cdot\big)$. Decompose the vector x below into orthogonal components with respect to W, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

(Notice that $\mathsf{w}_1 \not\perp \mathsf{w}_2$.)

6.4.4.- Given the matrix A below, find a basis for the space $R(\mathsf{A})^\perp$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}.$$

6.4.5.- Consider the subspace

$$W = \text{Span}\Big\{\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}\Big\}$$

in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

(a) Find a basis for the space $W^\perp$, that is, find a basis for the orthogonal complement of the space W.

(b) Use Theorem 6.4.1 to transform the basis of $W^\perp$ found in part (a) into an orthogonal basis.

6.4.6.- Consider the inner product space $\big(\mathbb{R}^4, \cdot\big)$, and find a basis for the orthogonal complement of the subspace W given by

$$W = \text{Span}\Big(\Big\{\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \ \begin{bmatrix} 2 \\ 4 \\ 1 \\ 6 \end{bmatrix}\Big\}\Big).$$

6.4.7.- Let X and Y be subspaces of a finite dimensional inner product space $\big(V, \langle\,,\rangle\big)$. Prove the following:

(a) $X \subset Y \quad \Rightarrow \quad Y^\perp \subset X^\perp$;

(b) $(X + Y)^\perp = X^\perp \cap Y^\perp$;

(c) Use part (b) to show that

$$(X \cap Y)^\perp = X^\perp + Y^\perp.$$

6.5. Gram-Schmidt method

We now describe the Gram-Schmidt orthogonalization method, which is a method to transform a linearly independent set of vectors into an orthonormal set. The method is based on projecting the i-th vector in the set onto the subspace spanned by the previous $(i-1)$ vectors.

Theorem 6.5.1 (Gram-Schmidt). *Let $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p\}$ be a linearly independent set in an inner product space $\big(V, \langle\,,\rangle\big)$. If the set $Y = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_p\}$ is defined as follows,*

$$\begin{aligned}
\boldsymbol{y}_1 &= \boldsymbol{x}_1, \\
\boldsymbol{y}_2 &= \boldsymbol{x}_2 - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle}{\|\boldsymbol{y}_1\|^2}\,\boldsymbol{y}_1, \\
&\vdots \\
\boldsymbol{y}_p &= \boldsymbol{x}_p - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_p\rangle}{\|\boldsymbol{y}_1\|^2}\,\boldsymbol{y}_1 - \cdots - \frac{\langle \boldsymbol{y}_{(p-1)}, \boldsymbol{x}_p\rangle}{\|\boldsymbol{y}_{(p-1)}\|^2}\,\boldsymbol{y}_{(p-1)},
\end{aligned}$$

then Y is an orthogonal set with $\mathrm{Span}(Y) = \mathrm{Span}(X)$. Furthermore, the set

$$Z = \left\{ \boldsymbol{z}_1 = \frac{\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|}, \cdots, \boldsymbol{z}_p = \frac{\boldsymbol{y}_p}{\|\boldsymbol{y}_p\|} \right\}$$

is an orthonormal set with $\mathrm{Span}(Z) = \mathrm{Span}(Y)$.

Remark: Using the notation in Theorem 6.4.2 we can write $\boldsymbol{y}_2 = \boldsymbol{x}_{2_\perp}$, where the projection is onto the subspace $\mathrm{Span}(\{\boldsymbol{y}_1\})$. Analogously, $\boldsymbol{y}_i = \boldsymbol{x}_{i_\perp}$, for $i = 2, \cdots, p$, where the projection is onto the subspace $\mathrm{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1}\})$.

Proof of Theorem 6.5.1: We first show that Y is an orthogonal set. It is simple to see that $\boldsymbol{y}_2 \in \mathrm{Span}(\{\boldsymbol{y}_1\})^\perp$, since

$$\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle = \langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle}{\|\boldsymbol{y}_1\|^2}\langle \boldsymbol{y}_1, \boldsymbol{y}_1\rangle = 0.$$

Assume that $\boldsymbol{y}_i \in \mathrm{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1}\})^\perp$, we then show that $\boldsymbol{y}_{i+1} \in \mathrm{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i\})^\perp$. Indeed, for $j = 1, \cdots, i$ holds

$$\langle \boldsymbol{y}_j, \boldsymbol{y}_{i+1}\rangle = \langle \boldsymbol{y}_j, \boldsymbol{x}_{i+1}\rangle - \frac{\langle \boldsymbol{y}_j, \boldsymbol{x}_{i+1}\rangle}{\|\boldsymbol{y}_j\|^2}\langle \boldsymbol{y}_j, \boldsymbol{y}_j\rangle = 0,$$

where we used that $\boldsymbol{y}_j \in \mathrm{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1}\})^\perp$, for all $j = 1, \cdots, i$. Therefore, Y is an orthogonal set (and so, a linearly independent set).

The proof that $\mathrm{Span}(X) = \mathrm{Span}(Y)$ has two steps: On the one hand, the elements in Y are linear combinations of elements in X, hence $\mathrm{Span}(Y) \subset \mathrm{Span}(X)$; on the other hand $\dim \mathrm{Span}(X) = \dim \mathrm{Span}(Y)$, since X and Y are both linearly independent sets with the same number of elements We conclude that $\mathrm{Span}(X) = \mathrm{Span}(Y)$. It is straightforward to see that Z is an orthonormal set, and since every element $\boldsymbol{z}_i \in Z$ is proportional to every $\boldsymbol{y}_i \in Y$, then $\mathrm{Span}(Y) = \mathrm{Span}(Z)$. This establishes the Theorem. □

Example 6.5.1: Use the Gram-Schmidt method to find an orthonormal basis for the inner product space $\big(\mathbb{R}^3, \cdot\big)$ from the ordered basis

$$X = \left(\mathsf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathsf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \mathsf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right).$$

SOLUTION: We first find an orthogonal basis. The first element is

$$y_1 = x_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \|y_1\|^2 = 2.$$

The second element is

$$y_2 = x_2 - \frac{y_1 \cdot x_2}{\|y_1\|^2}\, y_1,$$

where

$$y_1 \cdot x_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = 3.$$

A simple calculation shows

$$y_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

therefore,

$$y_2 = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \|y_2\|^2 = \frac{1}{2}.$$

Finally, the last element is

$$y_3 = x_3 - \frac{y_1 \cdot x_3}{\|y_1\|^2}\, y_1, - \frac{y_2 \cdot x_3}{\|y_2\|^2}\, y_2,$$

where

$$y_1 \cdot x_3 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 2, \qquad y_2 \cdot x_3 = \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0.$$

Another simple calculation shows

$$y_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

therefore,

$$y_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \|y_3\|^2 = 1.$$

The set $Y = \{y_1, y_2, y_3\}$ is an orthogonal set, while an orthonormal set is given by

$$Z = \left\{ z_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \; z_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \; z_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

◁

EXAMPLE 6.5.2: Consider the vector space $\mathbb{P}_3([-1, 1])$ with the inner product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} \boldsymbol{p}(x)\, \boldsymbol{q}(x)\, dx.$$

Given the basis $\{\boldsymbol{p}_0 = 1, \; \boldsymbol{p}_1 = x, \; \boldsymbol{p}_2 = x^2, \; \boldsymbol{p}_3 = x^3\}$, use the Gram-Schmidt method starting with the vector $\boldsymbol{p}_0$ to find an orthogonal basis for $\mathbb{P}_3([-1, 1])$.

SOLUTION: The first element in the new basis is

$$\boldsymbol{q}_0 = \boldsymbol{p}_0 = 1 \quad \Rightarrow \quad \|\boldsymbol{q}_0\|^2 = \int_{-1}^{1} dx = 2.$$

The second element is

$$\boldsymbol{q}_1 = \boldsymbol{p}_1 - \frac{\langle \boldsymbol{q}_0, \boldsymbol{p}_1 \rangle}{\|\boldsymbol{q}_0\|^2}\, \boldsymbol{q}_0.$$

It is simple to see that

$$\langle \boldsymbol{q}_0, \boldsymbol{p}_1 \rangle = \int_{-1}^{1} x\, dx = \frac{1}{2}x^2\Big|_{-1}^{1} = 0.$$

So we conclude that

$$\boldsymbol{q}_1 = \boldsymbol{p}_1 = x \quad \Rightarrow \quad \|\boldsymbol{q}_1\|^2 = \int_{-1}^{1} x^2\, dx = \frac{1}{3}x^3\Big|_{-1}^{1} \quad \Rightarrow \quad \|\boldsymbol{q}_1\|^2 = \frac{2}{3}.$$

The third element in the basis is

$$\boldsymbol{q}_2 = \boldsymbol{p}_2 - \frac{\langle \boldsymbol{q}_0, \boldsymbol{p}_2 \rangle}{\|\boldsymbol{q}_0\|^2}\, \boldsymbol{q}_0 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle}{\|\boldsymbol{q}_1\|^2}\, \boldsymbol{q}_1.$$

It is simple to see that

$$\langle \boldsymbol{q}_0, \boldsymbol{p}_2 \rangle = \int_{-1}^{1} x^2\, dx = \frac{1}{3}x^3\Big|_{-1}^{1} = \frac{2}{3},$$

$$\langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle = \int_{-1}^{1} x^3\, dx = \frac{1}{4}x^4\Big|_{-1}^{1} = 0.$$

Hence we obtain

$$\boldsymbol{q}_2 = \boldsymbol{p}_2 - \frac{2}{3}\frac{1}{2}\boldsymbol{q}_0 = x^2 - \frac{1}{3} \quad \Rightarrow \quad \boldsymbol{q}_2 = \frac{1}{3}\,(3x^2 - 1).$$

The norm square of this vector is

$$\begin{aligned} \|\boldsymbol{q}_2\|^2 &= \frac{1}{9}\int_{-1}^{1} (3x^2 - 1)(3x^2 - 1)\, dx \\ &= \frac{1}{9}\int_{-1}^{1} (9x^4 - 6x^2 + 1)\, dx \\ &= \frac{1}{9}\Big(\frac{9}{5}x^5 - 2x^3 + x\Big)\Big|_{-1}^{1} \\ &= \frac{8}{45}. \end{aligned}$$

Finally, the fourth vector of the orthogonal basis is given by

$$\boldsymbol{q}_3 = \boldsymbol{p}_3 - \frac{\langle \boldsymbol{q}_0, \boldsymbol{p}_3 \rangle}{\|\boldsymbol{q}_0\|^2}\, \boldsymbol{q}_0 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_3 \rangle}{\|\boldsymbol{q}_1\|^2}\, \boldsymbol{q}_1 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_3 \rangle}{\|\boldsymbol{q}_2\|^2}\, \boldsymbol{q}_2.$$

It is simple to see that

$$\langle \boldsymbol{q}_0, \boldsymbol{p}_3 \rangle = \int_{-1}^{1} x^3\, dx = \frac{1}{4}x^4\Big|_{-1}^{1} = 0,$$

$$\langle \boldsymbol{q}_1, \boldsymbol{p}_3 \rangle = \int_{-1}^{1} x^4\, dx = \frac{1}{5}x^5\Big|_{-1}^{1} = \frac{2}{5},$$

$$\langle \boldsymbol{q}_2, \boldsymbol{p}_3 \rangle = \frac{1}{3}\int_{-1}^{1} (3x^2 - 1)\, x^3\, dx = \frac{1}{3}\Big(\frac{1}{2}x^6 - \frac{1}{4}x^4\Big)\Big|_{-1}^{1} = 0.$$

Hence we obtain

$$\boldsymbol{q}_3 = \boldsymbol{p}_3 - \frac{2}{5}\frac{3}{2}\boldsymbol{q}_1 = x^3 - \frac{3}{5}x \quad \Rightarrow \quad \boldsymbol{q}_3 = \frac{1}{5}(5x^3 - 3x).$$

The orthogonal basis is then given by

$$\left\{\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = x,\ \boldsymbol{q}_2 = \frac{1}{3}\,(3x^2 - 1),\ \boldsymbol{q}_3 = \frac{1}{5}(5x^3 - 3x)\right\}.$$

These polynomials are proportional to the first three Legendre polynomials. The Legendre polynomials form an orthogonal set in the space $\mathbb{P}_\infty([-1, 1])$ of polynomials of all degrees. They play an important role in physics, since Legendre polynomials are solution of a particular differential equation that often appears in physics. ◁

6.5.1. **Exercises.**

6.5.1.- Find an orthonormal basis for the subspace of $\mathbb{R}^3$ spanned by the vectors

$$\Big\{\mathsf{u}_1 = \begin{bmatrix} -2 \\ 2 \\ -1 \end{bmatrix},\ \mathsf{u}_2 = \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}\Big\},$$

using the Gram-Schmidt process starting with the vector u_1.

6.5.2.- Let $W \subset \mathbb{R}^3$ be the subspace

$$\text{Span}\Big\{\mathsf{u}_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix},\ \mathsf{u}_2 = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}\Big\}.$$

(a) Find an orthonormal basis for W using the Gram-Schmidt method starting with the vector u_1.

(b) Decompose the vector x below as

$$\mathsf{x} = \mathsf{x}_{\shortparallel} + \mathsf{x}_{\perp},$$

with $\mathsf{x}_{\shortparallel} \in W$ and $\mathsf{x}_{\perp} \in W^{\perp}$, where

$$\mathsf{x} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}.$$

6.5.3.- Knowing that the column vectors in matrix A below form a linearly independent set, use the Gram-Schmidt method to find an orthonormal basis for $R(\mathsf{A})$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 5 \\ 0 & 2 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

6.5.4.- Use the Gram-Schmidt method to find an orthonormal basis for $R(\mathsf{A})$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 1 & 0 & 3 \end{bmatrix}.$$

6.5.5.- Consider the vector space $\mathbb{P}_2([0,1])$ with inner product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\,dx.$$

Use the Gram-Schmidt method on the ordered basis

$$\big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2\big),$$

starting with vector $\boldsymbol{p}_0$, to obtain an orthogonal basis for $\mathbb{P}_2([0,1])$.

6.6. The adjoint operator

6.6.1. **The Riesz Representation Theorem.** The Riesz Representation Theorem is a statement concerning linear functionals on an inner product space. Recall that a *linear functional* on a vector space V over a scalar field $\mathbb{F}$ is a linear function $f : V \to \mathbb{F}$, that is, for all $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and all a, $b \in \mathbb{F}$ holds $f(a\boldsymbol{x} + b\boldsymbol{y}) = a\, f(\boldsymbol{x}) + b\, f(\boldsymbol{y}) \in \mathbb{F}$. An example of a linear functional on $\mathbb{R}^3$ is the function

$$\mathbb{R}^3 \ni \mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto f(\mathsf{x}) = 3x_1 + 2x_2 + x_3 \in \mathbb{R}.$$

This function can be expressed in terms of the dot product in $\mathbb{R}^3$ as follows

$$f(\mathsf{x}) = \mathsf{u} \cdot \mathsf{x}, \qquad \mathsf{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

The Riesz Representation Theorem says that what we did in this example can be done in the general case. In an inner product space $\big(V, \langle\, , \rangle\big)$ every linear functional f can be expressed in terms of the inner product.

Theorem 6.6.1. *Consider a finite dimensional inner product space $\big(V, \langle\, , \rangle\big)$ over the scalar field $\mathbb{F}$. For every linear functional $f : V \to \mathbb{F}$ there exists a unique vector $\boldsymbol{u}_f \in V$ such that holds*

$$f(\boldsymbol{v}) = \langle \boldsymbol{u}_f, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in V.$$

Proof of Theorem 9.4.1: Introduce the set

$$N = \{\, \boldsymbol{v} \in V \;:\; f(\boldsymbol{v}) = 0 \,\} \subset V.$$

This set is the analogous to linear functionals of the null space of linear operators. Since f is a linear function the set N is a subspace of V. (Proof: Given two elements $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in N$ and two scalars a, $b \in \mathbb{F}$, holds that $f(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) = a\, f(\boldsymbol{v}_1) + b\, f(\boldsymbol{v}_2) = 0 + 0$, so $(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) \in N$.) Introduce the orthogonal complement of N, that is,

$$N^\perp = \{\boldsymbol{w} \in V \;:\; \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 0 \;\forall \boldsymbol{v} \in V\},$$

which is also a subspace of V. If $N^\perp = \{\boldsymbol{0}\}$, then $N = \big(N^\perp\big)^\perp = \big(\{\boldsymbol{0}\}\big)^\perp = V$. Since the null space of f is the whole vector space, the functional f is identically zero, so only for the choice $\boldsymbol{u}_f = \boldsymbol{0}$ holds $f(\boldsymbol{v}) = \langle \boldsymbol{0}, \boldsymbol{v} \rangle$ for all $\boldsymbol{v} \in V$.

In the case that $N^\perp \neq \{\boldsymbol{0}\}$ we now show that this space cannot be very big, in fact it has dimension one, as the following argument shows. Choose $\tilde{\boldsymbol{u}} \in N^\perp$ such that $f(\tilde{\boldsymbol{u}}) = 1$. Then notice that for every $\boldsymbol{w} \in N^\perp$ the vector $\boldsymbol{w} - f(\boldsymbol{w})\tilde{\boldsymbol{u}}$ is trivially in $N^\perp$ but it is also in N, since

$$f\big(\boldsymbol{w} - f(\boldsymbol{w})\, \tilde{\boldsymbol{u}}\big) = f(\boldsymbol{w}) - f(\boldsymbol{w})\, f(\tilde{\boldsymbol{u}}) = f(\boldsymbol{w}) - f(\boldsymbol{w}) = 0.$$

A vector both in N and $N^\perp$ must vanish, so $\boldsymbol{w} = f(\boldsymbol{w})\, \tilde{\boldsymbol{u}}$. Then every vector in $N^\perp$ is proportional to $\tilde{\boldsymbol{u}}$, so $\dim N^\perp = 1$. This information is used to split any vector $\boldsymbol{v} \in V$ as follows $\boldsymbol{v} = a\, \tilde{\boldsymbol{u}} + \boldsymbol{x}$ where $\boldsymbol{x} \in V$ and $a \in \mathbb{F}$. It is clear that

$$f(\boldsymbol{v}) = f(a\, \tilde{\boldsymbol{u}} + \boldsymbol{x}) = a\, f(\tilde{\boldsymbol{u}}) + f(\boldsymbol{x}) = a\, f(\tilde{\boldsymbol{u}}) = a.$$

However, the function with values $g(\boldsymbol{v}) = \left\langle \dfrac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, \boldsymbol{v} \right\rangle$ has precisely the same values as f, since for all $\boldsymbol{v} \in V$ holds

$$g(\boldsymbol{v}) = \left\langle \frac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, \boldsymbol{v} \right\rangle = \left\langle \frac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, (a\, \tilde{\boldsymbol{u}} + \boldsymbol{x}) \right\rangle = \frac{a}{\|\tilde{\boldsymbol{u}}\|^2} \langle \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}} \rangle + \frac{1}{\|\tilde{\boldsymbol{u}}\|^2} \langle \tilde{\boldsymbol{u}}, \boldsymbol{x} \rangle = a.$$

Therefore, choosing $\boldsymbol{u}_f = \tilde{\boldsymbol{u}}/\|\tilde{\boldsymbol{u}}\|^2$, holds that

$$f(\boldsymbol{v}) = \langle \boldsymbol{u}_f, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in V.$$

Since $\dim N^\perp = 1$, the choice of $\boldsymbol{u}_f$ is unique. This establishes the Theorem. □

6.6.2. **The adjoint operator.** Given a linear operator defined on an inner product space, a new linear operator can be defined through an equation involving the inner product.

Proposition 6.6.2. *If $\boldsymbol{T} \in L(V)$ is a linear operator on a finite-dimensional inner product space $\big(V, \langle\,,\,\rangle\big)$, then there exists a unique linear operator $\boldsymbol{T}^* \in L(V)$, called the* **adjoint** *of $\boldsymbol{T}$, such that for all vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$ holds*

$$\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u}) \rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u} \rangle.$$

Proof of Proposition 9.4.2: We first establish the following statement: For every vector $\boldsymbol{u} \in V$ there exists a unique vector $\boldsymbol{w} \in V$ such that

$$\langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u} \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle \qquad \forall \boldsymbol{v} \in V. \tag{6.18}$$

The proof starts noticing that for a fixed $\boldsymbol{u} \in V$ the scalar-valued function $f_u : V \to \mathbb{F}$ given by $f_u(\boldsymbol{v}) = \langle u, \boldsymbol{T}(\boldsymbol{v}) \rangle$ is a linear functional. Therefore, the Riesz Representation Theorem 6.6.1 implies that there exists a unique vector $\boldsymbol{w} \in V$ such that $f_u(\boldsymbol{v}) = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$. This establishes that for every vector $\boldsymbol{u} \in V$ there exists a unique vector $\boldsymbol{w} \in V$ such that Eq. (6.18) holds. Now that this statement is proven we can define a map, that we choose to denote as $\boldsymbol{T}^* : V \to V$, given by $\boldsymbol{u} \mapsto \boldsymbol{T}^*(\boldsymbol{u}) = \boldsymbol{w}$. We now show that this map $\boldsymbol{T}^*$ is linear. Indeed, for all $\boldsymbol{u}_1$, $\boldsymbol{u}_2 \in V$ and all a, $b \in \mathbb{F}$ holds

$$\begin{aligned}
\langle \boldsymbol{v}, \boldsymbol{T}^*(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) \rangle &= \langle \boldsymbol{T}(\boldsymbol{v}), (a\boldsymbol{u}_1 + b\boldsymbol{u}_2) \rangle \qquad \forall \boldsymbol{v} \in V, \\
&= a \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}_1 \rangle + b \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}_2 \rangle \\
&= a \langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u}_1) \rangle + b \langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u}_2) \rangle \\
&= \big\langle \boldsymbol{v}, \big[a\, \boldsymbol{T}^*(\boldsymbol{u}_1) + b\, \boldsymbol{T}^*(\boldsymbol{u}_2) \big] \big\rangle \qquad \forall \boldsymbol{v} \in V,
\end{aligned}$$

hence $\boldsymbol{T}^*(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) = a\, \boldsymbol{T}^*(\boldsymbol{u}_1) + b\, \boldsymbol{T}^*(\boldsymbol{u}_2)$. This establishes the Proposition. □

The next result relates the adjoint of a linear operator with the concept of the adjoint of a square matrix introduced in Sect. 2.2. Recall that given a basis in the vector space, every linear operator has associated a unique square matrix. Let us use the notation $[\boldsymbol{T}]$ and $[\boldsymbol{T}^*]$ for the matrices on a given basis of the operators $\boldsymbol{T}$ and $\boldsymbol{T}^*$, respectively.

Proposition 6.6.3. *Let $\big(V, \langle\,,\,\rangle\big)$ be a finite-dimensional vector space, let $\mathcal{V}$ be an orthonormal basis of V, and let $[\boldsymbol{T}]$ be the matrix of the linear operator $\boldsymbol{T} \in L(V)$ in the basis $\mathcal{V}$. Then, the matrix of the adjoint operator $\boldsymbol{T}^*$ in the basis $\mathcal{V}$ is given by $[\boldsymbol{T}^*] = [\boldsymbol{T}]^*$.*

Proposition 6.6.3 says that the matrix of the adjoint operator is the adjoint of the matrix of the operator, however this is true only in the case that the basis used to compute the respective matrices is orthonormal. If the basis is not orthonormal, the relation between the matrices $[\boldsymbol{T}]$ and $[\boldsymbol{T}^*]$ is more involved.

Proof of Proposition 6.6.3: Let $\mathcal{V} = \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ be an orthonormal basis of V, that is,

$$\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j. \end{cases}$$

The components of two arbitrary vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$ in the basis $\mathcal{V}$ is denoted as follows

$$\boldsymbol{u} = \sum_i u_i \boldsymbol{e}_i, \qquad \boldsymbol{v} = \sum_i v_i \boldsymbol{e}_i.$$

The action of the operator $\boldsymbol{T}$ can also be decomposed in the basis $\mathcal{V}$ as follows

$$\boldsymbol{T}(\boldsymbol{e}_j) = \sum_i [\boldsymbol{T}]_{ij}\, \boldsymbol{e}_i, \qquad [\boldsymbol{T}]_{ij} = [\boldsymbol{T}(\boldsymbol{e}_j)]_i.$$

We use the same notation for the adjoint operator, that is,

$$\boldsymbol{T}^*(\boldsymbol{e}_j) = \sum_i [\boldsymbol{T}^*]_{ij}\, \boldsymbol{e}_i, \qquad [\boldsymbol{T}^*]_{ij} = [\boldsymbol{T}^*(\boldsymbol{e}_j)]_i.$$

The adjoint operator is defined such that the equation $\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u})\rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}\rangle$ holds for all $\boldsymbol{u}$, $\boldsymbol{v} \in V$. This equation can be expressed in terms of components in the basis $\mathcal{V}$ as follows

$$\sum_{ijk} \big\langle v_i \boldsymbol{e}_i, u_j [\boldsymbol{T}^*(\boldsymbol{e}_j)]_k \boldsymbol{e}_k \big\rangle = \sum_{ijk} \big\langle v_i [\boldsymbol{T}(\boldsymbol{e}_i)]_k \boldsymbol{e}_k, u_j \boldsymbol{e}_j \big\rangle,$$

that is,

$$\sum_{ijk} \overline{v}_i u_j [\boldsymbol{T}^*]_{kj} \langle \boldsymbol{e}_i, \boldsymbol{e}_k\rangle = \sum_{ijk} \overline{v}_i \overline{[\boldsymbol{T}]}_{ki} u_j \langle \boldsymbol{e}_k, \boldsymbol{e}_j\rangle.$$

Since the basis $\mathcal{V}$ is orthonormal we obtain the equation

$$\sum_{ij} \overline{v}_i u_j [\boldsymbol{T}^*]_{ij} = \sum_{ijk} \overline{v}_i \overline{[\boldsymbol{T}]}_{ji} u_j,$$

which holds for all vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$, so we conclude

$$[\boldsymbol{T}^*]_{ij} = \overline{[\boldsymbol{T}]}_{ji} \quad \Leftrightarrow \quad [\boldsymbol{T}^*] = \overline{[\boldsymbol{T}]}^{\,T} \quad \Leftrightarrow \quad [\boldsymbol{T}^*] = [\boldsymbol{T}]^*.$$

This establishes the Proposition. □

EXAMPLE 6.6.1: Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$. Find the adjoint of the linear operator $\boldsymbol{T}$ with matrix in the standard basis of $\mathbb{C}^3$ given by

$$[\boldsymbol{T}(\boldsymbol{x})] = \begin{bmatrix} x_1 + 2ix_2 + ix_3 \\ ix_1 - x_3 \\ x_1 - x_2 + ix_3 \end{bmatrix}, \qquad [\boldsymbol{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

SOLUTION: The matrix of this operator in the standard basis of $\mathbb{C}^3$ is given by

$$[\boldsymbol{T}] = \begin{bmatrix} 1 & 2i & i \\ i & 0 & -1 \\ 1 & -1 & i \end{bmatrix}.$$

Since the standard basis is an orthonormal basis with respect to the dot product, Proposition 6.6.3 implies that

$$[\boldsymbol{T}^*] = [\boldsymbol{T}]^* = \begin{bmatrix} 1 & 2i & i \\ i & 0 & -1 \\ 1 & -1 & i \end{bmatrix}^* = \begin{bmatrix} 1 & -i & 1 \\ -2i & 0 & -1 \\ -i & -1 & -i \end{bmatrix} \quad \Rightarrow \quad [\boldsymbol{T}^*(\boldsymbol{x})] = \begin{bmatrix} x_1 - ix_2 + x_3 \\ -2ix_1 - x_3 \\ -ix_1 - x_2 - ix_3 \end{bmatrix}.$$

◁

6.6.3. **Normal operators.** Recall now that the commutator of two linear operators $\boldsymbol{T}$, $\boldsymbol{S} \in L(V)$ is the linear operator $[\boldsymbol{T}, \boldsymbol{S}] \in L(V)$ given by

$$[\boldsymbol{T}, \boldsymbol{S}](\boldsymbol{u}) = \boldsymbol{T}(\boldsymbol{S}(\boldsymbol{u})) - \boldsymbol{S}(\boldsymbol{T}(\boldsymbol{u})) \qquad \forall \boldsymbol{u} \in V.$$

Two operators $\boldsymbol{T}$, $\boldsymbol{S} \in L(V)$ are said to commute iff their commutator vanishes, that is, $[\boldsymbol{T}, \boldsymbol{S}] = \boldsymbol{0}$. Examples of operators that commute are two rotations on the plane. Examples of operators that do not commute are two arbitrary rotations in space.

Definition 6.6.4. *A linear operator $\boldsymbol{T}$ defined on a finite-dimensional inner product space $\big(V, \langle\,,\rangle\big)$ is called a* **normal operator** *iff holds $[\boldsymbol{T}, \boldsymbol{T}^*] = \boldsymbol{0}$, that is, the operator commutes with its adjoint.*

An interesting characterization of normal operators is the following: A linear operator $\boldsymbol{T}$ on an inner product space is normal iff $\|\boldsymbol{T}(\boldsymbol{u})\| = \|\boldsymbol{T}^*(\boldsymbol{u})\|$ holds for all $\boldsymbol{u} \in V$. Normal operators are particularly important because for these operators hold the Spectral Theorem, which we study in Chapter 9.

Two particular cases of normal operators are often used in physics. A linear operator $\boldsymbol{T}$ on an inner product space is called a **unitary operator** iff $\boldsymbol{T}^* = \boldsymbol{T}^{-1}$, that is, the adjoint is the inverse operator. Unitary operators are normal operators, since

$$\boldsymbol{T}^* = \boldsymbol{T}^{-1} \quad \Rightarrow \quad \begin{cases} \boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{I}, \\ \boldsymbol{T}^*\boldsymbol{T} = \boldsymbol{I}, \end{cases} \quad \Rightarrow \quad [\boldsymbol{T}, \boldsymbol{T}^*] = \boldsymbol{0}.$$

Unitary operators preserve the length of a vector, since

$$\|\boldsymbol{v}\|^2 = \langle \boldsymbol{v}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{T}^{-1}(\boldsymbol{T}(\boldsymbol{v}))\rangle = \langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{T}(\boldsymbol{v}))\rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{T}(\boldsymbol{v})\rangle = \|\boldsymbol{T}(\boldsymbol{v})\|^2.$$

Unitary operators defined on a complex inner product space are particularly important in quantum mechanics. The particular case of unitary operators on a real inner product space are called **orthogonal operators**. So, orthogonal operators do not change the length of a vector. Examples of orthogonal operators are rotations in $\mathbb{R}^3$.

A linear operator $\boldsymbol{T}$ on an inner product space is called an **Hermitian operator** iff $\boldsymbol{T}^* = \boldsymbol{T}$, that is, the adjoint is the original operator. This definition agrees with the definition of Hermitian matrices given in Chapter 2.

EXAMPLE 6.6.2: Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$ and the linear operator $\boldsymbol{T}$ with matrix in the standard basis of $\mathbb{C}^3$ given by

$$[\boldsymbol{T}(\boldsymbol{x})] = \begin{bmatrix} x_1 - ix_2 + x_3 \\ ix_1 - x_3 \\ x_1 - x_2 + x_3 \end{bmatrix}, \qquad [\boldsymbol{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Show that $\boldsymbol{T}$ is Hermitian.

SOLUTION: We need to compute the adjoint of $\boldsymbol{T}$. The matrix of this operator in the standard basis of $\mathbb{C}^3$ is given by

$$[\boldsymbol{T}] = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Since the standard basis is an orthonormal basis with respect to the dot product, Proposition 6.6.3 implies that

$$[\boldsymbol{T}^*] = [\boldsymbol{T}]^* = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}^* = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} = [\boldsymbol{T}].$$

Therefore, $\boldsymbol{T}^* = \boldsymbol{T}$. ◁

6.6.4. **Bilinear forms.**

Definition 6.6.5. *A* **bilinear form** *on a vector space V over $\mathbb{F}$ is a function $a : V \times V \to \mathbb{F}$ linear on both arguments, that is, for all $\boldsymbol{u}$, $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in V$ and all b_1, $b_2 \in \mathbb{F}$ holds*

$$a\big(\boldsymbol{u}, (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2)\big) = b_1 a(\boldsymbol{u}, \boldsymbol{v}_1) + b_2 a(\boldsymbol{u}, \boldsymbol{v}_2),$$
$$a\big((b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2), \boldsymbol{u}\big) = b_1 a(\boldsymbol{v}_1, \boldsymbol{u}) + b_2 a(\boldsymbol{v}_2, \boldsymbol{u}).$$

The bilinear form $a : V \times V \to \mathbb{F}$ is called **symmetric** *iff for all $\boldsymbol{u},\ \boldsymbol{v} \in V$ holds*

$$a(\boldsymbol{u}, \boldsymbol{v}) = a(\boldsymbol{v}, \boldsymbol{u}).$$

An example of a symmetric bilinear form is any the inner product on a real vector space. Indeed, given a real inner product space $\big(V, \langle\,,\rangle\big)$, the function $\langle\,,\rangle : V \times V \to \mathbb{R}$ is a symmetric bilinear form, since it is symmetric and

$$\langle \boldsymbol{u}, (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2)\rangle = b_1\langle \boldsymbol{u}, \boldsymbol{v}_1\rangle + b_2\langle \boldsymbol{u}, \boldsymbol{v}_2\rangle.$$

We will shortly see that another example of a bilinear form appears on the weak formulation of the boundary value problem in Eq. (7.30).

On the other hand, an inner product on a complex vector space is not a bilinear form, since it is conjugate linear on the first argument instead of linear. Such functions are called sesquilinear forms. That is, a **sesquilinear form** on a complex vector space V is a function $a : V \times V \to \mathbb{C}$ that is conjugate linear on the first argument and linear on the second argument. Sesquilinear forms are important in the case of studying differential equations involving complex functions.

Definition 6.6.6. *Consider a bilinear form $a : V \times V \to \mathbb{F}$ on an inner product space $\big(V, \langle\,,\rangle\big)$ over $\mathbb{F}$. The bilinear form a is called* **positive** *iff there exists a real number $k > 0$ such that for all $\boldsymbol{u} \in V$ holds*

$$a(\boldsymbol{u}, \boldsymbol{u}) \geqslant k\,\|\boldsymbol{u}\|^2.$$

The bilinear form a is called **bounded** *iff there exists a real number $K > 0$ such that for all $\boldsymbol{u},\ \boldsymbol{v} \in V$ holds*

$$a(\boldsymbol{u}, \boldsymbol{v}) \leqslant K\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|.$$

An example of a positive bilinear form is any inner product on a real vector space. The Schwarz inequality implies that such inner product is also a bounded bilinear form. In fact, an inner product on a real vector space is a symmetric, positive, bounded bilinear form. We will shortly see that the bilinear form that appears on a weak formulation of the boundary value problem in Eq. (7.30) is symmetric, positive and bounded. We will see that these properties imply the existence and uniqueness of solutions to the weak formulation of the boundary value problem.

6.6.5. **Exercises.**

6.6.1.- .

6.6.2.- .

Chapter 7. Approximation methods

7.1. Best approximation

The first half of this Section is dedicated to show that the Fourier series approximation of a function is a particular case of the orthogonal decomposition of a vector onto a subspace in an inner product space, studied in Section 6.4. Once we realize that it is not difficult to see why such Fourier approximations are useful. We show that the orthogonal projection $\boldsymbol{x}_{\shortparallel}$ of a vector $\boldsymbol{x}$ onto a subspace U is the vector in the subspace U closest to $\boldsymbol{x}$. This is the origin of the name "best approximation" for $\boldsymbol{x}_{\shortparallel}$. What is intuitively clear in $\big(\mathbb{R}^3, \cdot\big)$ is true in every inner product space, hence it is true for the Fourier series approximation of a function. In the second half of this Section we show a deep relation between the Null space of a matrix and the Range space of the adjoint matrix. The former is the orthogonal complement of the latter. A consequence of this relation is a simple proof to the property that a matrix and its adjoint matrix have the same rank. Another consequence is given in the next Section, where we obtain a simple equation, called the normal equation, to find a least squares solution of an inconsistent linear system.

7.1.1. **Fourier expansions.** We have seen that orthonormal bases have a practical advantage over arbitrary basis. The components $[\boldsymbol{x}]_u$ of a vector $\boldsymbol{x}$ in an orthonormal basis $\mathcal{U}_n = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ of the inner product space $\big(V, \langle\,,\rangle\big)$ are given by the simple expression

$$[\boldsymbol{x}]_u = \begin{bmatrix} \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle \\ \vdots \\ \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{x} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle\, \boldsymbol{u}_n.$$

In the case that an orthonormal set $\mathcal{U}_p$ is not a basis of V, that is, $p < \dim V$, one can always introduce the orthogonal projection of a vector $\boldsymbol{x} \in V$ onto the subspace $U_p = \text{Span}(\mathcal{U}_p)$,

$$\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_p, \boldsymbol{x}\rangle\, \boldsymbol{u}_p.$$

We have seen that in this case, $\boldsymbol{x}_{\shortparallel} \neq \boldsymbol{x}$. In fact we called the difference vector $\boldsymbol{x}_{\perp} = \boldsymbol{x} - \boldsymbol{x}_{\shortparallel}$, since this vector satisfies that $\boldsymbol{x}_{\perp} \perp \boldsymbol{x}_{\shortparallel}$. We now give the projection vector $\boldsymbol{x}_{\shortparallel}$ a new name.

Definition 7.1.1. *The* ***Fourier expansion*** *of a vector $\boldsymbol{x} \in V$ with respect to an orthonormal set $\mathcal{U}_p = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\} \subset \big(V, \langle\,,\rangle\big)$ is the unique vector $\boldsymbol{x}_{\shortparallel} \in \text{Span}(\mathcal{U}_p)$ given by*

$$\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_p, \boldsymbol{x}\rangle\, \boldsymbol{u}_p. \tag{7.1}$$

The scalars $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$ are the ***Fourier coefficient*** *of the vector $\boldsymbol{x}$ with respect to the set $\mathcal{U}_p$.*

A reason to the name "Fourier expansion" to the orthogonal projection of a vector onto a subspace is given in the following example.

EXAMPLE 7.1.1: Given the vector space of continuous functions $V = C\big([-\ell, \ell], \mathbb{R}\big)$ with inner product given by $\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \displaystyle\int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx$, find the Fourier expansion of an arbitrary function $\boldsymbol{f} \in V$ with respect to the orthonormal set

$$\mathcal{U}_N = \Big\{ \boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}},\ \boldsymbol{u}_n = \frac{1}{\sqrt{\ell}} \cos\Big(\frac{n\pi x}{\ell}\Big),\ \boldsymbol{v}_n = \frac{1}{\sqrt{\ell}} \sin\Big(\frac{n\pi x}{\ell}\Big) \Big\}_{n=1}^{N}.$$

SOLUTION: Using Eq. (7.1) on a function $\boldsymbol{f} \in V$ we get

$$\boldsymbol{f}_{\shortparallel} = \langle \boldsymbol{u}_0, \boldsymbol{f}\rangle\, \boldsymbol{u}_0 + \sum_{n=1}^{N} \Big[\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle\, \boldsymbol{u}_n + \langle \boldsymbol{v}_n, \boldsymbol{f}\rangle\, \boldsymbol{v}_n \Big].$$

Introduce the vectors in $\mathcal{U}_N$ explicitly in the expression above,

$$\boldsymbol{f}_{\shortparallel} = \frac{1}{\sqrt{2\ell}}\langle \boldsymbol{u}_0, \boldsymbol{f}\rangle + \sum_{n=1}^{N}\Big[\frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle\,\cos\Big(\frac{n\pi x}{\ell}\Big) + \frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{v}_n, \boldsymbol{f}\rangle\,\sin\Big(\frac{n\pi x}{\ell}\Big)\Big].$$

Denoting by

$$a_0 = \frac{1}{\sqrt{2\ell}}\,\langle \boldsymbol{u}_0, \boldsymbol{f}\rangle \quad a_n = \frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle \quad b_n = \frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle,$$

then we get that

$$\boldsymbol{f}_{\shortparallel}(x) = a_0 + \sum_{n=1}^{N}\Big[a_n\,\cos\Big(\frac{n\pi x}{\ell}\Big) + b_n\,\sin\Big(\frac{n\pi x}{\ell}\Big)\Big],$$

where the coefficients a_0, a_n and b_n, for $n = 1, \cdots, N$ are the usual Fourier coefficients,

$$a_0 = \frac{1}{2\ell}\int_{-\ell}^{\ell}\boldsymbol{f}(x)\,dx, \quad a_n = \frac{1}{\ell}\int_{-\ell}^{\ell}\boldsymbol{f}(x)\cos\Big(\frac{n\pi x}{\ell}\Big)\,dx, \quad b_n = \frac{1}{\ell}\int_{-\ell}^{\ell}\boldsymbol{f}(x)\sin\Big(\frac{n\pi x}{\ell}\Big)\,dx.$$

◁

The example above is a good reason to name Fourier expansion the orthogonal projection of a vector onto a subspace. We already know that the Fourier expansion $\boldsymbol{x}_{\shortparallel}$ of a vector $\boldsymbol{x}$ with respect to a an orthonormal set $\mathcal{U}_p$ has a particular property, that is, $(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) = \boldsymbol{x}_{\perp} \in U_p^{\perp}$. This property means that $\boldsymbol{x}_{\shortparallel}$ is the best approximation of the vector $\boldsymbol{x}$ from within the subspace $\text{Span}(\mathcal{U}_p)$. See Fig. 45 for the case $V = \mathbb{R}^3$, with $\langle\,,\rangle = \cdot$, and $\text{Span}(\mathcal{U}_2)$ two-dimensional. We highlight this property in the following result.

Theorem 7.1.2 (Best approximation). *The Fourier expansion $\boldsymbol{x}_{\shortparallel}$ of a vector $\boldsymbol{x}$ with respect to an orthonormal set $\mathcal{U}_p$ in an inner product space, is the unique vector in the subspace $\text{Span}(\mathcal{U}_p)$ that is closest to $\boldsymbol{x}$, in the sense that*

$$\|\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}\| < \|\boldsymbol{x} - \boldsymbol{y}\| \qquad \forall\ \boldsymbol{y} \in \text{Span}(\mathcal{U}_p) - \{\boldsymbol{x}_{\shortparallel}\}.$$

Proof of Theorem 7.1.2: Recall that $\boldsymbol{x}_{\perp} = \boldsymbol{x} - \boldsymbol{x}_{\shortparallel}$ is orthogonal to $\text{Span}(U)$, that is $(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) \perp (\boldsymbol{x}_{\shortparallel} - \boldsymbol{y})$ for all $\boldsymbol{y} \in \text{Span}(U)$. Hence,

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) + (\boldsymbol{x}_{\shortparallel} - \boldsymbol{y})\|^2 = \|\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}\|^2 + \|\boldsymbol{x}_{\shortparallel} - \boldsymbol{y}\|^2, \tag{7.2}$$

where the las equality comes from Pythagoras Theorem. Eq. (7.2) says that $\|\boldsymbol{x} - \boldsymbol{y}\|$ is the smallest iff $\boldsymbol{y} = \boldsymbol{x}_{\shortparallel}$ and the smallest value is $\|\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}\|$. This establishes the Theorem. □

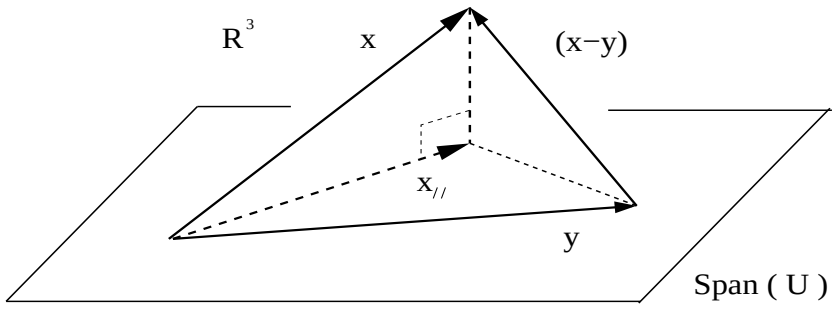

FIGURE 45. The Fourier expansion $\mathsf{x}_{\shortparallel}$ of the vector $\mathsf{x} \in \mathbb{R}^3$ is the best approximation of x from within $\text{Span}(U)$.

EXAMPLE 7.1.2: Given the vector space of continuous functions $V = C\big([-1,1],\mathbb{R}\big)$ with the inner product given by $\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \displaystyle\int_{-1}^{1} \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx$, find the Fourier expansion of the function $\boldsymbol{f}(x) = x$ with respect to the orthonormal set

$$\mathcal{U}_N = \Big\{ \boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}},\ \boldsymbol{u}_n = \frac{1}{\sqrt{\ell}}\cos\Big(\frac{n\pi x}{\ell}\Big),\ \boldsymbol{v}_n = \frac{1}{\sqrt{\ell}}\sin\Big(\frac{n\pi x}{\ell}\Big)\Big\}_{n=1}^{N}.$$

SOLUTION: We use the formulas in Example 7.1.1 above to the function $\boldsymbol{f}(x) = x$ on the interval $[-1,1]$, since $\ell = 1$. The coefficient a_0 above is given by

$$a_0 = \frac{1}{2}\int_{-1}^{1} x\,dx = \frac{1}{4}\big(x^2\big|_{-1}^{1}\big) \quad \Rightarrow \quad a_0 = 0.$$

The coefficients a_n, b_n, for $n = 1, \cdots, N$ computed with one integration by parts,

$$\int x\cos(n\pi x)\,dx = \frac{x}{n\pi}\sin(n\pi x) + \frac{1}{n^2\pi^2}\cos(n\pi x),$$
$$\int x\sin(n\pi x)\,dx = -\frac{x}{n\pi}\cos(n\pi x) + \frac{1}{n^2\pi^2}\sin(n\pi x).$$

The coefficients a_n vanish, since

$$a_n = \int_{-1}^{1} x\cos(n\pi x)\,dx = \Big[\frac{x}{n\pi}\sin(n\pi x)\Big]\Big|_{-1}^{1} + \frac{1}{n^2\pi^2}\cos(n\pi x)\Big|_{-1}^{1} \quad \Rightarrow \quad a_n = 0.$$

The coefficients b_n are given by

$$b_n = \int_{-1}^{1} x\sin(n\pi x)\,dx = -\Big[\frac{x}{n\pi}\cos(n\pi x)\Big]\Big|_{-1}^{1} + \frac{1}{n^2\pi^2}\sin(n\pi x)\Big|_{-1}^{1} \quad \Rightarrow \quad b_n = \frac{2(-1)^{(n+1)}}{n\pi}.$$

Therefore, the Fourier expansion of $\boldsymbol{f}(x) = x$ with respect to $\mathcal{U}_N$ is given by

$$\boldsymbol{f}_{\shortparallel}(x) = \frac{2}{\pi}\sum_{n=1}^{N}\frac{(-1)^{(n+1)}}{n}\sin(n\pi x).$$

REMARK: First, a simpler proof that the coefficients a_0 and a_n vanish is to realized that we are integrating an odd function on the interval $[-1,1]$. The odd function is the product of an odd function times an even function. Second, Theorem 7.1.2 tells us that the function $\boldsymbol{f}_{\shortparallel}$ above is only combination of the sine and cosine functions in $\mathcal{U}_N$ that approximates best the function $\boldsymbol{f}(x) = x$ on the interval $[-1,1]$. ◁

7.1.2. **Null and range spaces of a matrix.** The null and range spaces associated with a matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ and its adjoint matrix A^* are deeply related.

Theorem 7.1.3. *For every matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ holds $N(\mathsf{A}) = R(\mathsf{A}^*)^{\perp}$ and $N(\mathsf{A}^*) = R(\mathsf{A})^{\perp}$.*

Since for every subspace W on a finite dimensional inner product space holds $(W^{\perp})^{\perp} = W$, we also have the relations

$$N(\mathsf{A})^{\perp} = R(\mathsf{A}^*), \qquad N(\mathsf{A}^*)^{\perp} = R(\mathsf{A}).$$

In the case of real-valued matrices, the Theorem above says that

$$N(\mathsf{A}) = R(\mathsf{A}^T)^{\perp} \quad \text{and} \quad N(\mathsf{A}^T) = R(\mathsf{A})^{\perp}.$$

Before we state the proof of this Theorem let us review the following notation: Given an $m \times n$ matrix $\mathsf{A} \in \mathbb{F}^{m,n}$, we write it either in terms of column vectors $\mathsf{A}_{:j} \in \mathbb{F}^m$ for $j = 1, \cdots, n$, or in terms of row vectors $\mathsf{A}_{i:} \in \mathbb{F}^n$ for $i = 1, \cdots, m$, as follows,

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{:1} & \cdots & \mathsf{A}_{:n} \end{bmatrix}, \qquad \mathsf{A} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{m:} \end{bmatrix}.$$

Since the same type of definition holds for the $n \times m$ matrix A^*, that is,

$$\mathsf{A}^* = \begin{bmatrix} (\mathsf{A}^*)_{:1} & \cdots & (\mathsf{A}^*)_{:m} \end{bmatrix}, \qquad \mathsf{A}^* = \begin{bmatrix} (\mathsf{A}^*)_{1:} \\ \vdots \\ (\mathsf{A}^*)_{n:} \end{bmatrix},$$

then we have the relations

$$(\mathsf{A}_{:j})^* = (\mathsf{A}^*)_{j:}, \qquad (\mathsf{A}_{i:})^* = (\mathsf{A}^*)_{:i}.$$

For example, consider the 2×3 matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \Rightarrow \quad \mathsf{A}_{:1} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \ \mathsf{A}_{:2} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \ \mathsf{A}_{:3} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \begin{array}{l} \mathsf{A}_{1:} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \\ \mathsf{A}_{2:} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}. \end{array}$$

The transpose is a 3×2 matrix that can be written as follows

$$\mathsf{A}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \quad \Rightarrow \quad (\mathsf{A}^T)_{:1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ (\mathsf{A}^T)_{:2} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \begin{array}{l} (\mathsf{A}^T)_{1:} = \begin{bmatrix} 1 & 4 \end{bmatrix}, \\ (\mathsf{A}^T)_{2:} = \begin{bmatrix} 2 & 5 \end{bmatrix}, \\ (\mathsf{A}^T)_{3:} = \begin{bmatrix} 3 & 6 \end{bmatrix}. \end{array}$$

So, for example we have the relation

$$(\mathsf{A}_{:3})^T = \begin{bmatrix} 3 & 6 \end{bmatrix} = (\mathsf{A}^T)_{3:}, \qquad (\mathsf{A}_{2:})^T = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = (\mathsf{A}^T)_{:2}.$$

Proof of Theorem 7.1.3: We first show that the $N(\mathsf{A}) = R(\mathsf{A}^*)^\perp$. A vector $\mathsf{x} \in \mathbb{F}^n$ belongs to $N(\mathsf{A})$ iff holds

$$\mathsf{A}\mathsf{x} = \mathsf{0} \quad \Leftrightarrow \quad \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{m:} \end{bmatrix} \mathsf{x} = \mathsf{0} \quad \Leftrightarrow \quad \begin{bmatrix} \big[(\mathsf{A}^*)_{:1}\big]^* \\ \vdots \\ \big[(\mathsf{A}^*)_{:m}\big]^* \end{bmatrix} \mathsf{x} = \mathsf{0} \quad \Leftrightarrow \quad \begin{cases} (\mathsf{A}^*)_{:1} \cdot \mathsf{x} = 0, \\ \qquad \vdots \\ (\mathsf{A}^*)_{:m} \cdot \mathsf{x} = 0. \end{cases}$$

So, $\mathsf{x} \in N(\mathsf{A})$ iff x is orthogonal to every column vector in A^*, that is, $\mathsf{x} \in R(\mathsf{A}^*)^\perp$.

The equation $N(\mathsf{A}^*) = R(\mathsf{A})^\perp$ comes from $N(\mathsf{B}) = R(\mathsf{B}^*)^\perp$ taking $\mathsf{B} = \mathsf{A}^*$. Nevertheless, we repeat the proof above, just to understand the previous argument. A vector $\mathsf{y} \in \mathbb{F}^m$ belongs to $N(\mathsf{A}^*)$ iff

$$\mathsf{A}^*\mathsf{y} = \mathsf{0} \quad \Leftrightarrow \quad \begin{bmatrix} (\mathsf{A}^*)_{1:} \\ \vdots \\ (\mathsf{A}^*)_{n:} \end{bmatrix} \mathsf{y} = \mathsf{0} \quad \Leftrightarrow \quad \begin{bmatrix} (\mathsf{A}_{:1})^* \\ \vdots \\ (\mathsf{A}_{:n})^* \end{bmatrix} \mathsf{y} = \mathsf{0} \quad \Leftrightarrow \quad \begin{cases} \mathsf{A}_{:1} \cdot \mathsf{y} = 0, \\ \qquad \vdots \\ \mathsf{A}_{:n} \cdot \mathsf{y} = 0. \end{cases}$$

So, $\mathsf{y} \in N(\mathsf{A}^*)$ iff y is orthogonal to every column vector in A, that is, $\mathsf{y} \in R(\mathsf{A})^\perp$. This establishes the Theorem. □

EXAMPLE 7.1.3: Verify Theorem 7.1.3 for the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$.

SOLUTION: We first find the $N(A)$, that is, all $\mathsf{x} \in \mathbb{R}^3$ solutions of $\mathsf{Ay} = \mathsf{0}$. Gauss operations on matrix A imply

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = x_3, \\ x_2 = -2x_3, \\ x_3 \text{ free}, \end{cases} \quad \Rightarrow \quad N(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \Big\} \Big).$$

It is simple to find $R(\mathsf{A}^T)$, since

$$R(\mathsf{A}^T) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \Big\} \Big).$$

Theorem 7.1.3 is verified, since

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 1-4+3 = 0, \quad \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 3-4+1 = 0 \quad \Rightarrow \quad N(\mathsf{A}) = R(\mathsf{A}^T)^\perp.$$

Let us verify the same Theorem for A^T. We first find $N(\mathsf{A}^T)$, that is, all $\mathsf{y} \in \mathbb{R}^2$ solutions of $\mathsf{A}^T\mathsf{y} = \mathsf{0}$. Gauss operations on matrix A^T imply

$$\begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad N(\mathsf{A}^T) = \{\mathsf{0}\}.$$

The space $R(\mathsf{A})$ is given by

$$R(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \Big\} \Big) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix} \Big\} \Big) = \mathbb{R}^2.$$

Since $(\mathbb{R}^2)^\perp = \{\mathsf{0}\}$, Theorem 7.1.3 is verified. ◁

Theorem 7.1.3 provides a simple proof for a result we used in Chapter 2.

Theorem 7.1.4. *For every matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ holds that* $\text{rank}(\mathsf{A}) = \text{rank}(\mathsf{A}^*)$.

Proof of Theorem 7.1.4: Recall the Nullity-Rank result in Corollary 5.1.8, which says that for all matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ holds $\dim N(\mathsf{A}) + \dim R(\mathsf{A}) = n$. Equivalently,

$$\dim R(\mathsf{A}) = n - \dim N(\mathsf{A}) = n - \dim R(\mathsf{A}^*)^\perp,$$

since $N(\mathsf{A}) = R(\mathsf{A}^*)^\perp$. From the orthogonal decomposition $\mathbb{F}^n = R(\mathsf{A}^*) \oplus R(\mathsf{A}^*)^\perp$ we know that $\dim R(\mathsf{A}^*) = n - \dim R(\mathsf{A}^*)^\perp$. We then conclude that

$$\dim R(\mathsf{A}) = \dim R(\mathsf{A}^*).$$

This establishes the Theorem. □

7.1.3. **Exercises.**

7.1.1.- Consider the inner product space $\big(\mathbb{R}^3, \cdot\,\big)$ and the orthonormal set $\mathcal{U}$,

$$\Big\{ \mathsf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \; \mathsf{u}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Big\}.$$

Find the best approximation of x below in the subspace $\mathrm{Span}(\mathcal{U})$, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

7.1.2.- Consider the inner product space $\big(\mathbb{R}^{2,2}, \langle\,,\,\rangle_F\big)$ and the orthonormal set $\mathcal{U} = \{\mathsf{E}_1, \mathsf{E}_2\}$, where

$$\mathsf{E}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \; \mathsf{E}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Find the best approximation of matrix A below in the subspace $\mathrm{Span}(\mathcal{U})$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

7.1.3.- Consider the inner product space $\mathbb{P}_2([0,1])$, with $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\,dx$, and the subspace $U = \mathrm{Span}(\mathcal{U})$, where $\mathcal{U} = \{\boldsymbol{q}_0 = 1, \;\; \boldsymbol{q}_1 = (x - \frac{1}{2})\}$.

(a) Show that $\mathcal{U}$ is an orthogonal set.

(b) Find $\boldsymbol{r}_{\shortparallel}$, the best approximation with respect to U of the polynomial $\boldsymbol{r}(x) = 2x + 3x^2$.

(c) Verify whether $(\boldsymbol{r} - \boldsymbol{r}_{\shortparallel}) \in U^{\perp}$ or not.

7.1.4.- Consider the space $C^{\infty}([-\ell, \ell], \mathbb{R})$ with inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx,$$

and the orthonormal set $\mathcal{U}$ given by

$$\boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}}$$
$$\boldsymbol{u}_1 = \frac{1}{\sqrt{\ell}} \cos\Big(\frac{\pi x}{\ell}\Big)$$
$$\boldsymbol{v}_1 = \frac{1}{\sqrt{\ell}} \sin\Big(\frac{\pi x}{\ell}\Big).$$

Find the best approximation of

$$\boldsymbol{f}(x) = \begin{cases} x & 0 \leqslant x \leqslant \ell, \\ -x & -\ell \leqslant x < 0. \end{cases}$$

in the space $\mathrm{Span}(\mathcal{U})$

7.1.5.- For the matrix $\mathsf{A} \in \mathbb{R}^{3,3}$ below, verify that $N(\mathsf{A}) = R(\mathsf{A}^T)^{\perp}$ and that $N(\mathsf{A}^T) = R(\mathsf{A})^{\perp}$, where

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 1 \\ -1 & -1 & 0 \\ -2 & -1 & -1 \end{bmatrix}.$$

7.2. Least squares

7.2.1. **The normal equation.** We describe the least squares method to find approximate solutions to inconsistent linear systems. The method is often used to find the best parameters that fit experimental data. The parameters are the unknowns of the linear system, and the experimental data determines the matrix of coefficients and the source vector of the system. Such a linear system usually contains more equations than unknowns, and it is inconsistent, since there are no parameters that fit all the data exactly. We start introducing the notion of least squares solution of a possibly inconsistent linear system.

Definition 7.2.1. *Given a matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *and a vector* $\mathsf{b} \in \big(\mathbb{F}^m, \cdot\big)$*, the vector* $\hat{\mathsf{x}} \in \mathbb{F}^n$ *is called a* ***least squares solution*** *of the linear system* $\mathsf{Ax} = \mathsf{b}$ *iff holds*

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\| \leqslant \|\mathsf{y} - \mathsf{b}\| \qquad \forall\, \mathsf{y} \in R(\mathsf{A}).$$

The problem we study is to find the least squares solution to an $m \times n$ linear system $\mathsf{Ax} = \mathsf{b}$. In the case that $\mathsf{b} \in R(\mathsf{A})$ the linear system $\mathsf{Ax} = \mathsf{b}$ is consistent and the least squares solution $\hat{\mathsf{x}}$ is the actual solution of the system, hence $\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\| = 0$. In the case that b does not belong to $R(\mathsf{A})$, the linear system $\mathsf{Ax} = \mathsf{b}$ is inconsistent. In such a case the least squares solution $\hat{\mathsf{x}}$ is the vector in $\mathbb{R}^n$ with the property that $\mathsf{A}\hat{\mathsf{x}}$ is a vector in $R(\mathsf{A})$ closest to b in the inner product space $\big(\mathbb{R}^m, \cdot\big)$. A sketch of this situation for a matrix $\mathsf{A} \in \mathbb{R}^{3,2}$ is given in Fig. 46.

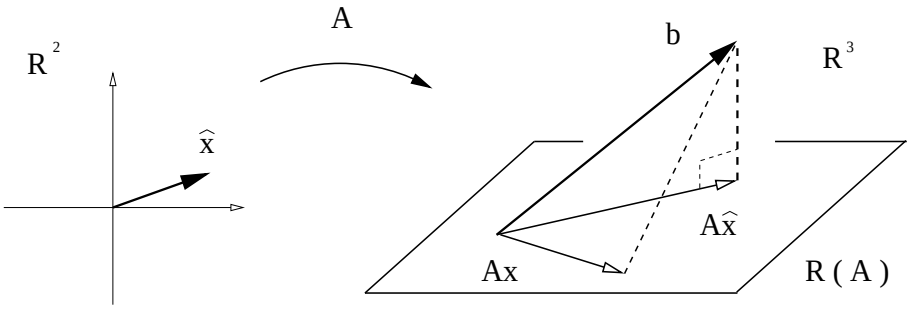

Figure 46. The meaning of the least squares solution $\hat{\mathsf{x}} \in \mathbb{R}^2$ for the 3×2 inconsistent linear system $\mathsf{Ax} = \mathsf{b}$ is that the vector $\mathsf{A}\hat{\mathsf{x}}$ is the closest to b in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

The solution to the problem of finding a least squares solution to a linear system is summarized in the following result.

Theorem 7.2.2. *Given a matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *and a vector* b *in the inner product space* $\big(\mathbb{F}^m, \cdot\big)$*, the vector* $\hat{\mathsf{x}} \in \mathbb{F}^n$ *is a least squares solution of the* $m \times n$ *linear system* $\mathsf{Ax} = \mathsf{b}$ *iff* $\hat{\mathsf{x}}$ *is solution to the* $n \times n$ *linear system, called* ***normal equation****,*

$$\mathsf{A}^*\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}. \tag{7.3}$$

Furthermore, the least squares solution $\hat{\mathsf{x}}$ *is unique iff the column vectors of matrix* A *form a linearly independent set.*

Remark: In the case that $\mathbb{F} = \mathbb{R}$, the normal equation reduces to $\mathsf{A}^T\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^T\mathsf{b}$.

Proof of Theorem 7.2.2: We are interested in finding a vector $\hat{\mathsf{x}} \in \mathbb{F}^n$ such that $\mathsf{A}\hat{\mathsf{x}}$ is the best approximation in $R(\mathsf{A})$ of vector $\mathsf{b} \in \mathbb{F}^m$. That is, we want to find $\hat{\mathsf{x}} \in \mathbb{F}^n$ such that

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\| \leqslant \|\mathsf{y} - \mathsf{b}\| \qquad \forall\, \mathsf{y} \in R(\mathsf{A}).$$

Theorem 7.1.2 says that the best approximation of b is when $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}_{\shortparallel}$, where $\mathsf{b}_{\shortparallel}$ is the orthogonal projection of b onto the subspace $R(\mathsf{A})$. This means that

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) \in R(\mathsf{A})^{\perp} = N(\mathsf{A}^*) \quad \Leftrightarrow \quad \mathsf{A}^*(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) = 0.$$

We then conclude that $\hat{\mathsf{x}}$ must be solution of the normal equation

$$\mathsf{A}^*\mathsf{A}\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}.$$

The furthermore can be shown as follows. The column vectors of matrix A form a linearly independent set iff $N(\mathsf{A}) = \{0\}$. Lemma 7.2.3 stated below establishes that, for all matrix A holds that $N(\mathsf{A}) = N(\mathsf{A}^*\mathsf{A})$. This result in our case implies that $N(\mathsf{A}^*\mathsf{A}) = \{0\}$. Since matrix $\mathsf{A}^*\mathsf{A}$ is a square, $n \times n$, matrix, we conclude that it is invertible. This is equivalent to say that the solution $\hat{\mathsf{x}}$ to the normal equation is unique; moreover, it is given by $\hat{\mathsf{x}} = (\mathsf{A}^*\mathsf{A})^{-1}\mathsf{A}^*\mathsf{b}$. This establishes the Theorem. □

In the proof of Theorem 7.2.2 above we used the following result:

Lemma 7.2.3. *If $\mathsf{A} \in \mathbb{F}^{m,n}$, then $N(\mathsf{A}) = N(\mathsf{A}^*\mathsf{A})$.*

Proof of Lemma 7.2.3: We first show that $N(\mathsf{A}) \subset N(\mathsf{A}^*\mathsf{A})$. Indeed,

$$\mathsf{x} \in N(\mathsf{A}) \quad \Rightarrow \quad \mathsf{A}\mathsf{x} = 0 \quad \Rightarrow \quad \mathsf{A}^*\mathsf{A}\mathsf{x} = 0 \quad \Rightarrow \quad \mathsf{x} \in N(\mathsf{A}^*\mathsf{A}).$$

Now, suppose that there exists $\mathsf{x} \in N(\mathsf{A}^*\mathsf{A})$ such that $\mathsf{x} \notin N(\mathsf{A})$. Therefore, $\mathsf{x} \neq 0$ and $\mathsf{A}\mathsf{x} \neq 0$, which imply that

$$0 \neq \|\mathsf{A}\mathsf{x}\|^2 = \mathsf{x}^*\mathsf{A}^*\mathsf{A}\mathsf{x} \quad \Rightarrow \quad \mathsf{A}^*\mathsf{A}\mathsf{x} \neq 0.$$

However, this last equation contradicts the assumption that $\mathsf{x} \in N(\mathsf{A}^*\mathsf{A})$. Therefore, we conclude that $N(\mathsf{A}) = N(\mathsf{A}^*\mathsf{A})$. This establishes the Lemma. □

EXAMPLE 7.2.1: Show that the 3×2 linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$ is inconsistent; then find a least squares solutions $\hat{\mathsf{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$ to that system, where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}.$$

SOLUTION: We first show that the linear system above is inconsistent, since Gauss operation on the augmented matrix $[\mathsf{A}|\mathsf{b}]$ imply

$$\left[\begin{array}{cc|c} 1 & 3 & -1 \\ 2 & 2 & 1 \\ 3 & 1 & -1 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & -4 & 3 \\ 0 & -8 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & -4 & 3 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{array}\right].$$

In order to find the least squares solution to the system above we first construct the normal equation. We need to compute

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix}, \qquad \mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}.$$

Therefore, the normal equation is given by

$$\begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}.$$

Since the column vectors of A form a linearly independent set, matrix $\mathsf{A}^T\mathsf{A}$ is invertible,

$$\left(\mathsf{A}^T\mathsf{A}\right)^{-1} = \frac{1}{96} \begin{bmatrix} 14 & -10 \\ -10 & 14 \end{bmatrix} = \frac{1}{48} \begin{bmatrix} 7 & -5 \\ -5 & 7 \end{bmatrix}.$$

The least squares solution is unique and given by

$$\hat{x} = \frac{1}{24}\begin{bmatrix} 7 & -5 \\ -5 & 7 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \hat{x} = \frac{1}{12}\begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

REMARK: We now verify that $(A\hat{x} - b) \in R(A)^{\perp}$. Indeed,

$$A\hat{x} - b = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix}\frac{1}{12}\begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} = -\frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad A\hat{x} - b = \frac{2}{3}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

Since

$$\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 - 4 + 3 = 0, \qquad \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = 3 - 4 + 1 = 0,$$

we have verified that $(A\hat{x} - b) \in R(A)^{\perp}$. ◁

We finish this Subsection with an alternative proof of Theorem 7.2.2 in the particular case that involves real-valued matrices, that is, $\mathbb{F} = \mathbb{R}$. The proof is interesting in its own, since it is based in solving a constrained minimization problem.

Alternative proof of Theorem 7.2.2 for $\mathbb{F} = \mathbb{R}$: The vector $\hat{x} \in \mathbb{R}^n$ is a least squares solution of the system $Ax = b$ iff the function $f : \mathbb{R}^n \to R$ given by $f(x) = \|A\hat{x} - b\|^2$ has a minimum at $x = \hat{x}$. We then find all minima of function f. We first express f as follows,

$$\begin{aligned} f(x) &= (Ax - b) \cdot (Ax - b) \\ &= (Ax) \cdot (Ax) - 2b \cdot (Ax) + b \cdot b \\ &= x^T A^T A x - 2b^T A x + b^T b. \end{aligned}$$

We now need to find all solutions to the equation $\nabla_x f(x) = 0$. Recalling the definition of a gradient vector

$$\nabla_x f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix},$$

it is simple to see that, for any vector $a \in \mathbb{R}^n$, holds,

$$\nabla_x(a^T x) = a, \qquad \nabla_x(x^T a) = a.$$

Therefore, the gradient of f is given by

$$\nabla_x f = 2A^T A x - 2A^T b.$$

We are interested in the stationary points, the $\hat{x}$ solutions of

$$\nabla_x f(\hat{x}) = 0 \quad \Leftrightarrow \quad A^T A\hat{x} = A^T b.$$

We conclude that all stationary points $\hat{x}$ are solutions of the normal equation, Eq. (7.3). These stationary points must be a minimum of f, since f is quadratic on the vector components x_i having the degree two terms all positive coefficients. This establishes the first part of Theorem 7.2.2 in the case that $\mathbb{F} = \mathbb{R}$. □

7.2.2. **Least squares fit.** It is often desirable to construct a mathematical model to describe the results of an experiment. This may involve fitting an algebraic curve to the given experimental data. The least squares method can be used to find the best parameters that fit the data.

EXAMPLE 7.2.2: (**Linear fit**) The simplest situation is the case where the best curve fitting the data is a straight line. More precisely, suppose that the result of an experiment is the following collection of ordered numbers

$$\big\{(t_1, b_1), \cdots, (t_m, b_m)\big\}, \qquad m \geqslant 2,$$

and suppose that a plot on a plane of the result of this experiment is given in Fig. 47. (For example, from measuring the vertical displacement b_i in a spring when a weight t_i is attached to it.) Find the best line $y(t) = \hat{x}_2\, t + \hat{x}_1$ that approximate these points in least squares sense. The latter means to find the numbers $\hat{x}_2$, $\hat{x}_1 \in \mathbb{R}$ such that $\sum_{i=1}^{m} |\Delta b_i|^2$ is the smallest possible, where

$$\Delta b_i = b_i - y(t_i) \quad \Leftrightarrow \quad \Delta b_i = b_i - (\hat{x}_2\, t_i + \hat{x}_1), \qquad i = 1, \cdots, m.$$

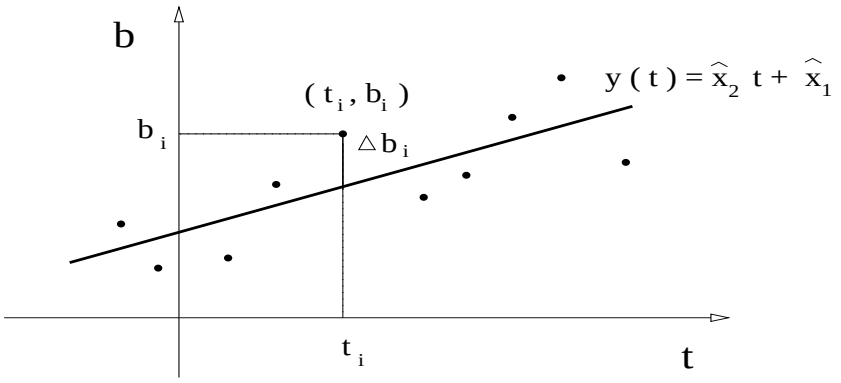

FIGURE 47. Sketch of the best line $y(t) = \hat{x}_2 t + \hat{x}_1$ fitting the set of points (t_i, b_i), for $i = 1, \cdots, 10$.

SOLUTION: Let us rewrite this problem as the least squares solution of an $m \times 2$ linear system, which in general is inconsistent. We are interested to find $\hat{x}_2$, $\hat{x}_1$ solution of the linear system

$$\begin{array}{c} y(t_1) = b_1 \\ \vdots \\ y(t_m) = b_m \end{array} \quad \Leftrightarrow \quad \begin{array}{c} \hat{x}_1 + t_1 \hat{x}_2 = b_1 \\ \vdots \\ \hat{x}_1 + t_m \hat{x}_2 = b_m \end{array} \quad \Leftrightarrow \quad \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Introducing the notation

$$\mathsf{A} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \qquad \hat{\mathsf{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

we are then interested in finding the solution $\hat{\mathsf{x}}$ of the $m\times 2$ linear system $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}$. Introducing also the vector

$$\Delta\mathsf{b} = \begin{bmatrix} \Delta b_1 \\ \vdots \\ \Delta b_m \end{bmatrix},$$

it is clear that $\mathsf{A}\hat{\mathsf{x}} - \mathsf{b} = \Delta\mathsf{b}$, and so we obtain the important relation

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2 = \|\Delta\mathsf{b}\|^2 = \sum_{i=1}^{m} \big(\Delta b_i\big)^2.$$

Therefore, the vector $\hat{\mathsf{x}}$ that minimizes the square of the deviation from the line, $\sum_{i=1}^{m}(\Delta b_i)^2$, is precisely the same vector $\hat{\mathsf{x}} \in \mathbb{R}^2$ that minimizes the number $\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2$. We studied the latter problem at the begining of this Section. We called it a least squares problem, and the solution $\hat{\mathsf{x}}$ is the solution of the normal equation

$$\mathsf{A}^T\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^T\mathsf{b}.$$

It is simple to see that

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix},$$

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Therefore, we are interested in finding the solution to the 2×2 linear system

$$\begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Suppose that at least one of the t_i is different from 1, then matrix $\mathsf{A}^T\mathsf{A}$ is invertible and the inverse is

$$\big(\mathsf{A}^T\mathsf{A}\big)^{-1} = \frac{1}{m\sum t_i^2 - \big(\sum t_i\big)^2} \begin{bmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{bmatrix}.$$

We conclude that the solution to the normal equation is

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{m\sum t_i^2 - \big(\sum t_i\big)^2} \begin{bmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{bmatrix} \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

So, the slope $\hat{x}_2$ and vertical intercept $\hat{x}_1$ of the best fitting line are given by

$$\hat{x}_2 = \frac{m\sum t_i b_i - (\sum t_i)(\sum b_i)}{m\sum t_i^2 - \big(\sum t_i\big)^2}, \qquad \hat{x}_1 = \frac{(\sum t_i^2)(\sum b_i) - (\sum t_i)(\sum t_i b_i)}{m\sum t_i^2 - \big(\sum t_i\big)^2}.$$

◁

EXAMPLE 7.2.3: (**Polynomial fit**) Find the best polynomial of degree $(n-1) \geqslant 0$, say $p(t) = \hat{x}_n\, t^{(n-1)} + \cdots + \hat{x}_1$, that approximates in least squares sense the set of points

$$\big\{(t_1, b_1), \cdots, (t_m, b_m)\big\}, \qquad m \geqslant n.$$

(See Fig. 48 for an example in the case that the fitting curve is a parabola, $n = 3$.) Following Example 7.2.2, the least squares approximation means to find the numbers $\hat{x}_n, \cdots, \hat{x}_1 \in \mathbb{R}$ such that $\sum_{i=1}^{m} |\Delta b_i|^2$ is the smallest possible, where

$$\Delta b_i = b_i - p(t_i) \quad \Leftrightarrow \quad \Delta b_i = b_i - (\hat{x}_n\, t_i^{(n-1)} + \cdots + \hat{x}_1), \qquad i = 1, \cdots, m.$$

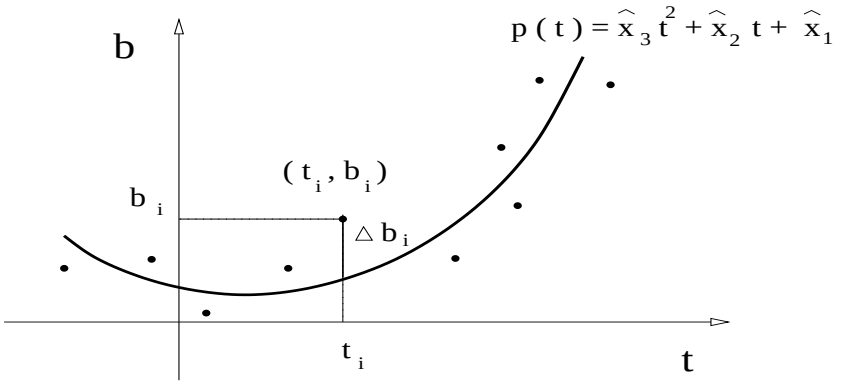

FIGURE 48. Sketch of the best parabola $p(t) = \hat{x}_3 t^2 + \hat{x}_2 t + \hat{x}_1$ fitting the set of points (t_i, b_i), for $i = 1, \cdots, 10$.

SOLUTION: We rewrite this problem as the least squares solution of an $m \times n$ linear system, which in general is inconsistent. We are interested to find $\hat{x}_n, \cdots, \hat{x}_1$ solution of the linear system

$$\begin{array}{c} p(t_1) = b_1 \\ \vdots \\ p(t_m) = b_m \end{array} \quad \Leftrightarrow \quad \begin{array}{c} \hat{x}_1 + \cdots + t_1^{(n-1)} \hat{x}_n = b_1 \\ \vdots \\ \hat{x}_1 + \cdots + t_m^{(n-1)} \hat{x}_n = b_m \end{array} \quad \Leftrightarrow \quad \begin{bmatrix} 1 & \cdots & t_1^{(n-1)} \\ \vdots & & \vdots \\ 1 & \cdots & t_m^{(n-1)} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Introducing the notation

$$\mathsf{A} = \begin{bmatrix} 1 & \cdots & t_1^{(n-1)} \\ \vdots & & \vdots \\ 1 & \cdots & t_m^{(n-1)} \end{bmatrix}, \qquad \hat{\mathsf{x}} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

we are then interested in finding the solution $\hat{\mathsf{x}}$ of the $m \times n$ linear system $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}$. Introducing also the vector

$$\Delta\mathsf{b} = \begin{bmatrix} \Delta b_1 \\ \vdots \\ \Delta b_m \end{bmatrix},$$

it is clear that $\mathsf{A}\hat{\mathsf{x}} - \mathsf{b} = \Delta\mathsf{b}$, and so we obtain the important relation

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2 = \|\Delta\mathsf{b}\|^2 = \sum_{i=1}^{m} \big(\Delta b_i\big)^2.$$

Therefore, the vector $\hat{\mathsf{x}}$ that minimizes the square of the deviation from the line, $\sum_{i=1}^{m} (\Delta b_i)^2$, is precisely the same vector $\hat{\mathsf{x}} \in \mathbb{R}^2$ that minimizes the number $\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2$. We studied the latter problem at the begining of this Section. We called it a least squares problem, and the solution $\hat{\mathsf{x}}$ is the solution of the normal equation

$$\mathsf{A}^T \mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^T \mathsf{b}. \tag{7.4}$$

It is simple to see that Eq. (7.4) is an $n \times n$ linear system, since

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ t_1^{(n-1)} & \cdots & t_m^{(n-1)} \end{bmatrix} \begin{bmatrix} 1 & \cdots & t_1^{(n-1)} \\ \vdots & & \vdots \\ 1 & \cdots & t_m^{(n-1)} \end{bmatrix},$$

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ t_1^{(n-1)} & \cdots & t_m^{(n-1)} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

We do not compute these expressions explicitly here. In the case that the columns of A form a linearly independent set, the solution $\hat{\mathsf{x}}$ to the normal equation is

$$\hat{\mathsf{x}} = \left(\mathsf{A}^T\mathsf{A}\right)^{-1}\mathsf{A}^T\mathsf{b}.$$

The components of $\hat{\mathsf{x}}$ provide the parameters for the best polynomial fitting the data in least squares sense. ◁

7.2.3. **Linear correlation.** In statistics a correlation coefficient measures the departure of two random variables from independence. For centered data, that is, for data with zero average, the correlation coefficient can be viewed as the cosine of the angle in an abstract $\mathbb{R}^n$ space between two vectors constructed with the random variables data. We now define and find the correlation coefficient for two variables as given in Example 7.2.2.

Once again, suppose that the result of an experiment is the following collection of ordered numbers

$$\big\{(t_1, b_1), \cdots, (t_m, b_m)\big\}, \qquad m \geqslant 2. \tag{7.5}$$

Introduce the vectors e, t, $\mathsf{b} \in \mathbb{R}^m$ as follows,

$$\mathsf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \mathsf{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Before introducing the correlation coefficient, let us use these vectors above to write down the least squares coefficients x found in Example 7.2.2. The matrix of coefficients can be written as $\mathsf{A} = [\mathsf{e}, \mathsf{t}]$, therefore,

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} \mathsf{e}^T \\ \mathsf{t}^T \end{bmatrix} [\mathsf{e}, \mathsf{t}] = \begin{bmatrix} \mathsf{e}\cdot\mathsf{e} & \mathsf{t}\cdot\mathsf{e} \\ \mathsf{e}\cdot\mathsf{t} & \mathsf{t}\cdot\mathsf{t} \end{bmatrix}, \qquad \mathsf{A}^T\mathsf{b} = \begin{bmatrix} \mathsf{e}^T \\ \mathsf{t}^T \end{bmatrix} \mathsf{b} = \begin{bmatrix} \mathsf{e}\cdot\mathsf{b} \\ \mathsf{t}\cdot\mathsf{b} \end{bmatrix}.$$

The least squares solution can be written as follows,

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{t}\cdot\mathsf{e})^2} \begin{bmatrix} \mathsf{t}\cdot\mathsf{t} & -\mathsf{e}\cdot\mathsf{t} \\ -\mathsf{e}\cdot\mathsf{t} & \mathsf{e}\cdot\mathsf{e} \end{bmatrix} \begin{bmatrix} \mathsf{e}\cdot\mathsf{b} \\ \mathsf{t}\cdot\mathsf{b} \end{bmatrix},$$

that is,

$$\hat{\mathsf{x}}_2 = \frac{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{b}) - (\mathsf{e}\cdot\mathsf{t})(\mathsf{e}\cdot\mathsf{b})}{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{t}\cdot\mathsf{e})^2}, \qquad \hat{x}_1 = \frac{(\mathsf{e}\cdot\mathsf{b})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{e}\cdot\mathsf{t})(\mathsf{t}\cdot\mathsf{b})}{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{t}\cdot\mathsf{e})^2}.$$

Introduce the average values

$$\bar{t} = \frac{\mathsf{e}\cdot\mathsf{t}}{\mathsf{e}\cdot\mathsf{e}}, \qquad \bar{b} = \frac{\mathsf{e}\cdot\mathsf{b}}{\mathsf{e}\cdot\mathsf{e}}.$$

These are indeed the average values of t_i and b_i, since

$$\mathsf{e}\cdot\mathsf{e} = m, \qquad \mathsf{e}\cdot\mathsf{t} = \sum_{i=1}^{m} t_i, \qquad \mathsf{e}\cdot\mathsf{b} = \sum_{i=1}^{m} b_i.$$

Introduce the zero-average vectors $\hat{\mathbf{t}} = (\mathbf{t} - \bar{t}\,\mathbf{e})$ and $\hat{\mathbf{b}} = (\mathbf{b} - \bar{b}\,\mathbf{e})$. The *correlation coefficient* of the data given in (7.5) is given by

$$\text{cor}(\mathbf{t}, \mathbf{b}) = \frac{\hat{\mathbf{t}} \cdot \hat{\mathbf{b}}}{\|\hat{\mathbf{t}}\|\,\|\hat{\mathbf{b}}\|}.$$

Therefore, the correlation coefficient between the data vectors $\mathbf{t}$ and $\mathbf{b}$ is the angle between the zero-average vectors $\hat{\mathbf{t}}$ and $\hat{\mathbf{b}}$ in $\mathbb{R}^m$.

In order to understand what measures this angle, let us consider the case where all the ordered pairs in (7.5) lies on a line, that is, there exists a solution $\hat{\mathbf{x}}$ of the linear system $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$ (a solution, not only a least squares solution). In that case we have

$$\hat{x}_1\mathbf{e} + \hat{x}_2\mathbf{t} = \mathbf{b} \quad \Rightarrow \quad \hat{x}_1 + \hat{x}_2\bar{t} = \bar{b},$$

and this implies that

$$\hat{x}_2\,(\mathbf{t} - \bar{t}\,\mathbf{e}) = (\mathbf{b} - \bar{b}\,\mathbf{e}) \quad \Leftrightarrow \quad \hat{x}_2\,\hat{\mathbf{t}} = \hat{\mathbf{b}} \quad \Leftrightarrow \quad \text{cor}(\mathbf{t}, \mathbf{b}) = 1.$$

That is, in the case that $\mathbf{t}$ is linearly related to $\mathbf{b}$ we obtain that the zero-average vectors $\hat{\mathbf{t}}$ and $\hat{\mathbf{b}}$ are parallel, so the correlation coefficient is equal one.

7.2.4. **QR-factorization.** The Gram-Schmidt method can be used to factor any $m \times n$ matrix $\mathbf{A}$ into a product of an $m \times n$ matrix $\mathbf{Q}$ with orthonormal column vectors and an upper triangular $n \times n$ matrix R. We will see that the QR-factorization is useful to solve the normal equation associated to a least squares problem.

Theorem 7.2.4. *If the column vectors of matrix $\mathbf{A} \in \mathbb{F}^{m,n}$ form a linearly independent set, then there exist matrices $\mathbf{Q} \in \mathbb{F}^{m,n}$ and $\mathbf{R} \in \mathbb{F}^{n,n}$ satisfying that $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}_n$, matrix $\mathbf{R}$ is upper triangular with positive diagonal elements, and the following equation holds*

$$\mathbf{A} = \mathbf{QR}.$$

Proof of Theorem 7.2.4: Use the Gram-Schmidt method to obtain an orthonormal set $\{\mathbf{q}_1, \cdots, \mathbf{q}_n\}$ from the column vectors of the $m \times n$ matrix $\mathbf{A} = [\mathbf{A}_{:1}, \cdots, \mathbf{A}_{:n}]$, that is,

$$\begin{aligned}
\mathbf{p}_1 &= \mathbf{A}_{:1} && \mathbf{q}_1 = \frac{\mathbf{p}_1}{\|\mathbf{p}_1\|}, \\
\mathbf{p}_2 &= \mathbf{A}_{:2} - (\mathbf{A}_{:2} \cdot \mathbf{q}_1)\,\mathbf{q}_1 && \mathbf{q}_2 = \frac{\mathbf{p}_2}{\|\mathbf{p}_2\|}, \\
&\vdots && \vdots \\
\mathbf{p}_n &= \mathbf{A}_{:n} - (\mathbf{A}_{:n} \cdot \mathbf{q}_1)\,\mathbf{q}_1 - \cdots - (\mathbf{A}_{:n} \cdot \mathbf{q}_{n-1})\,\mathbf{q}_{n-1} && \mathbf{q}_n = \frac{\mathbf{p}_n}{\|\mathbf{p}_n\|}.
\end{aligned}$$

Define matrix $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_n]$, which then satisfies the equation $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}_n$. Notice that the equations above can be expressed as follows,

$$\begin{aligned}
\mathbf{A}_{:1} &= \|\mathbf{p}_1\|\,\mathbf{q}_1, \\
\mathbf{A}_{:2} &= \|\mathbf{p}_2\|\,\mathbf{q}_2 + (\mathbf{q}_1 \cdot \mathbf{A}_{:2})\,\mathbf{q}_1 \\
&\vdots \\
\mathbf{A}_{:n} &= \|\mathbf{p}_n\|\,\mathbf{q}_n + (\mathbf{q}_1 \cdot \mathbf{A}_{:n})\,\mathbf{q}_1 + (\mathbf{q}_2 \cdot \mathbf{A}_{:n})\,\mathbf{q}_2 + \cdots + (\mathbf{q}_{n-1} \cdot \mathbf{A}_{:n})\,\mathbf{q}_{n-1}.
\end{aligned}$$

After some time staring at the equations above, one can rewrite it as a matrix product

$$\big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big] = \big[\mathsf{q}_1, \cdots, \mathsf{q}_n\big] \begin{bmatrix} \|\mathsf{p}_1\| & (\mathsf{q}_1 \cdot \mathsf{A}_{:2}) & \cdots & (\mathsf{q}_1 \cdot \mathsf{A}_{:n}) \\ 0 & \|\mathsf{p}_2\| & \cdots & (\mathsf{q}_2 \cdot \mathsf{A}_{:n}) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (\mathsf{q}_{n-1} \cdot \mathsf{A}_{:n}) \\ 0 & 0 & \cdots & \|\mathsf{p}_n\| \end{bmatrix} \tag{7.6}$$

Define matrix R by equation above as the matrix satisfying $\mathsf{A} = \mathsf{QR}$. Then, Eq. (7.6) says that matrix R is $n \times n$, upper triangular, with positive diagonal elements. This establishes the Theorem. □

EXAMPLE 7.2.4: Find the QR-factorization of matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

SOLUTION: First use the Gram-Schmidt method to transform the column vectors of matrix A into an orthonormal set. This was done in Example 6.5.1. The result defines the matrix Q as follows

$$\mathsf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Having matrix A and Q, and knowing that Theorem 7.2.4 is true, then we can compute matrix R by the equation $\mathsf{R} = \mathsf{Q}^T\mathsf{A}$. Since the column vectors of Q form an orthonormal set, we have that $\mathsf{Q}^T = \mathsf{Q}^{-1}$, and in this particular case $\mathsf{Q}^{-1} = \mathsf{Q}$, so matrix R is given by

$$\mathsf{R} = \mathsf{QA} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{R} = \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The QR-factorization of matrix A is then given by

$$\mathsf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

◁

The QR-factorization is useful to solve the normal equation in a least squares problem.

Theorem 7.2.5. *Assume that the matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ admits the QR-factorization $\mathsf{A} = \mathsf{QR}$. The vector $\hat{\mathsf{x}} \in \mathbb{F}^n$ is solution of the normal equation $\mathsf{A}^*\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}$ iff it is solution of*

$$\mathsf{R}\hat{\mathsf{x}} = \mathsf{Q}^*\mathsf{b}.$$

Proof of Theorem 7.2.5: Just introduce the QR-factorization into the normal equation $\mathsf{A}^*\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}$ as follows,

$$\big(\mathsf{R}^*\mathsf{Q}^*\big)\big(\mathsf{QR}\big)\,\hat{\mathsf{x}} = \mathsf{R}^*\mathsf{Q}^*\mathsf{b} \quad \Leftrightarrow \quad \mathsf{R}^*\mathsf{R}\,\hat{\mathsf{x}} = \mathsf{R}^*\mathsf{Q}^*\mathsf{b} \quad \Leftrightarrow \quad \mathsf{R}^*\big(\mathsf{R}\,\hat{\mathsf{x}} - \mathsf{Q}^*\mathsf{b}\big) = 0.$$

Since R is a square, upper triangular matrix with non-zero coefficients, we conclude that R is invertible. Therefore, from the last equation above we conclude that $\hat{\mathsf{x}}$ is solution of the normal equation iff holds

$$\mathsf{R}\hat{\mathsf{x}} = \mathsf{Q}^*\mathsf{b}.$$

This establishes the Theorem. □

7.2.5. **Exercises.**

7.2.1.- Consider the matrix A and the vector b given by

$$\mathsf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the linear system $\mathsf{Ax} = \mathsf{b}$.

(b) Verify that the solution $\hat{\mathsf{x}}$ satisfies

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) \in R(\mathsf{A})^{\perp}.$$

7.2.2.- Consider the matrix A and the vector b given by

$$\mathsf{A} = \begin{bmatrix} 2 & 2 \\ 0 & -1 \\ -2 & 0 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the linear system $\mathsf{Ax} = \mathsf{b}$.

(b) Find the orthogonal projection of the source vector b onto the subspace $R(\mathsf{A})$.

7.2.3.- Find all the least-squares solutions $\hat{\mathsf{x}}$ to the linear system $\mathsf{Ax} = \mathsf{b}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

7.2.4.- Find the best line in least-squares sense that fits the measurements, where t_1 is the independent variable and b_i is the dependent variable,

$$\begin{aligned} t_1 &= -2, & b_1 &= 4, \\ t_2 &= -1, & b_2 &= 3, \\ t_3 &= 0, & b_3 &= 1, \\ t_4 &= 2, & b_4 &= 0. \end{aligned}$$

7.2.5.- Find the correlation coefficient corresponding to the measurements given in Exercise **7.2.4** above.

7.2.6.- Use Gram-Schmidt method on the columns of matrix A below to find its QR factorization, where

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 2 & 1 \end{bmatrix}.$$

7.2.7.- Find the QR factorization of matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

7.3. Finite difference method

A differential equation is an equation where the unknown is a function and both the function itself and its derivatives appear in the equation. The differential equation is called linear iff the unknown function and its derivatives appear linearly in the equation. Solutions of linear differential equations can be approximated by solutions of appropriate $n \times n$ algebraic linear systems in the limit that n approaches infinity. The finite difference method is a way to obtain an $n \times n$ linear system from the original differential equation. *Derivatives are approximated by difference quotients, thus reducing a differential equation to an algebraic linear system.* Since a derivative can be approximated by infinitely many difference quotients, there are infinitely many $n \times n$ linear systems that approximate a differential equation. One tries to choose the linear system whose solution is the best approximation of the solution of the original differential equation. Computers are used to find the vector in $\mathbb{R}^n$ solution of the $n \times n$ linear system. Many approximations of the solution to the differential equation are obtained from this array of n numbers. One way to obtain a function from a vector in $\mathbb{R}^n$ is to find a degree n polynomial that contains all these n points. This is called a polynomial interpolation of the algebraic solution. In this Section we only show how to obtain $n \times n$ algebraic linear systems that approximate a simple differential equation.

7.3.1. **Differential equations.** A differential equation is an equation where the unknown is a function and both the function and its derivatives appear in the equation. A simple example is the following: Given a continuously differentiable function $f : [0, 1] \to \mathbb{R}$, find a function $u : [0, 1] \to \mathbb{R}$ solution of the differential equation

$$\frac{du}{dx}(x) = f(x),$$

To find a solution to a differential equation requires to perform appropriate integrations, thus integration constants are introduced in the solution. This suggests that the solution of a differential equation is not unique, and extra conditions must be added to the problem to select only one solution. In the differential equation above the solutions u are given by

$$u(x) = \int_0^x f(t)\, dt + c,$$

with $c \in \mathbb{R}$. An extra condition is needed to obtain a unique solution, for example the condition $u(0) = 1$. Then, the unique solution u is computed as follows

$$1 = u(0) = \int_0^0 f(t)\, dt + c \quad \Rightarrow \quad c = 1 \quad \Rightarrow \quad u(x) = \int_0^x f(t)\, dt + 1.$$

The example above is simple enough that no approximation is needed to obtain the solution.

An *ordinary differential equation* is a differential equation where the unknown function u depends on one variable, as in the example above. A *partial differential equation* is a differential equation where the unknown function depends on more than one variable and the equation contains derivatives of more than one variable. In this Section we use the finite difference method to find a solution to two different problems. The first one involves an ordinary differential equation while the second one involves a partial differential equation, called the heat equation.

The first problem is to find an approximate solution to a boundary value problem for an ordinary differential equation: Given a continuously differentiable function $f : [0, 1] \to \mathbb{R}$,

find a function $u : [0,1] \to \mathbb{R}$ solution of the boundary value problem

$$\frac{d^2u}{dx^2}(x) + \frac{du}{dx}(x) = f(x), \tag{7.7}$$

$$u(0) = u(1) = 0. \tag{7.8}$$

Finding the function u involves doing two integrations that introduce two integration constants. These constants are determined by the two conditions on $x = 0$ and $x = 1$ above, called boundary conditions, since they are conditions on the boundaries of the interval $[0,1]$. Because of this extra condition on the ordinary differential equation this problem is called a boundary value problem.

The second problem we study in this Section is to find an approximate solution to an initial-boundary value problem for the heat equation: Given the set $D = [0,\pi] \times [0,T]$, the positive constant κ, and infinitely differentiable functions $f : D \to \mathbb{R}$ and $g : [0,\pi] \to \mathbb{R}$, find the function $u : D \to \mathbb{R}$ solution of the problem

$$\frac{\partial u}{\partial t}(x,t) - \kappa \frac{\partial^2 u}{\partial x^2}(x,t) = f(x,t) \qquad (x,t) \in D, \tag{7.9}$$

$$u(0,t) = u(\pi,t) = 0 \qquad t \in [0,T], \tag{7.10}$$

$$u(x,0) = g(x) \qquad x \in [0,\pi]. \tag{7.11}$$

The partial differential equation in Eq. (7.9) is called the one-dimensional heat equation, since in the case that function u is the temperature of a material that depends on time t and one spatial direction x, the equation describes how the material temperature changes in time due to heat propagation. The positive constant κ is called the thermal diffusivity of the material. The condition given in Eq. (7.10) is called a boundary condition, since they are conditions on $x = 0$ and $x = \pi$ that hold for all $t \in [0,T]$, see Fig. 49. The condition given in Eq. (7.11) is called an initial condition, since it is a condition on the initial time $t = 0$ for all $x \in [0,\pi]$, see Fig. 49. Because of these two extra conditions on the partial differential equation this problem is called an initial-boundary value problem for the heat equation.

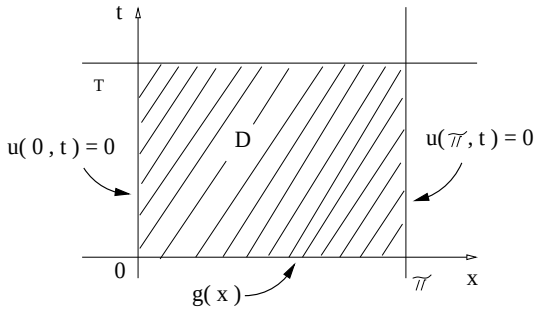

FIGURE 49. The domain $D = [0,\pi] \times [0,T]$ where the initial-boundary value problem for the heat equation is set up. We indicate the boundary data conditions $u(0,t) = 0$ and $u(\pi,t) = 0$, and the initial data function g.

7.3.2. **Difference quotients.** Finite difference methods transform a linear differential equation into an $n \times n$ algebraic linear system by replacing derivatives by difference quotients.

Derivatives can be approximated by difference quotients in many different ways. For example, a derivative of a function u can be expressed in the following equivalent ways,

$$\begin{aligned}\frac{du}{dx}(x) &= \lim_{\Delta x\to 0}\frac{u(x+\Delta x)-u(x)}{\Delta x},\\ &= \lim_{\Delta x\to 0}\frac{u(x)-u(x-\Delta x)}{\Delta x},\\ &= \lim_{\Delta x\to 0}\frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}.\end{aligned}$$

However, for a fixed nonzero value of Δx, the expressions below are, in general, different.

Definition 7.3.1. *The* ***forward difference quotient*** d_+ *and the* ***backward difference quotient*** d_- *of a continuous function $u:\mathbb{R}\to\mathbb{R}$ at $x\in\mathbb{R}$ are given by*

$$d_+u(x) = \frac{u(x+\Delta x)-u(x)}{\Delta x}, \qquad d_-u(x) = \frac{u(x)-u(x-\Delta x)}{\Delta x}.$$

The ***centered difference quotient*** d_c *of a continuous function u at $x\in\mathbb{R}$ is given by*

$$d_cu(x) = \frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}. \tag{7.12}$$

In the case that the function u has second continuous derivative, the Taylor Expansion Theorem implies that the forward and backward differences differ from the actual derivative by terms order Δx. The proof is not difficult, since

$$\begin{aligned}u(x+\Delta x) = u(x)+\frac{du}{dx}(x)\,\Delta x+O\big((\Delta x)^2\big) \quad &\Rightarrow \quad d_+u(x) = \frac{du}{dx}(x)+O\big(\Delta x\big),\\ u(x-\Delta x) = u(x)-\frac{du}{dx}(x)\,\Delta x+O\big((\Delta x)^2\big) \quad &\Rightarrow \quad d_-u(x) = \frac{du}{dx}(x)+O\big(\Delta x\big),\end{aligned}$$

where $O\big((\Delta x)^n\big)$ denotes a function satisfying $\big[O\big((\Delta x)^n\big)\big]/(\Delta x)^n$ approaches a constant as $\Delta x\to 0$. In the case that the function u has third continuous derivative, the Taylor Expansion Theorem implies that the centered difference quotient differs from the actual derivative by terms of order $(\Delta x)^2$. Again, the proof is not difficult, and it is based on the Taylor expansion of the function u. Compute this expansion in two different ways,

$$u(x+\Delta x) = u(x)+\frac{du}{dx}(x)\,\Delta x+\frac{1}{2}\,\frac{d^2u}{dx^2}(x)\,(\Delta x)^2+O\big((\Delta x)^3\big), \tag{7.13}$$

$$u(x-\Delta x) = u(x)-\frac{du}{dx}(x)\,\Delta x+\frac{1}{2}\,\frac{d^2u}{dx^2}(x)\,(\Delta x)^2+O\big((\Delta x)^3\big), \tag{7.14}$$

Subtracting the two expressions above we obtain that

$$u(x+\Delta x)-u(x-\Delta x) = 2\frac{du}{dx}(x)\,\Delta x+O\big((\Delta x)^3\big) \quad \Rightarrow \quad d_cu(x) = \frac{du}{dx}(x)+O\big((\Delta x)^2\big),$$

which establishes that centered difference quotients differ from the derivative by order $(\Delta x)^2$. If a function is infinitely continuously differentiable, centered difference quotients are more accurate than forward or backward differences.

Second and higher derivatives of a function can also be approximated by difference quotients. Again, there are infinitely many ways to approximate second derivatives by difference quotients. The freedom to choose difference quotients is higher for second derivatives than for first derivatives. We now present two difference quotients to give an idea of this freedom. On the one hand, one possible approximation is to use the centered difference quotient twice, since a second derivative is the derivative of the derivative function. Using the more precise notation

$$d_{c\Delta x}u(x) = \frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x},$$

it is not difficult to check that $\big(d_{c\Delta x}\big)^2 = d_{c\Delta x}(d_{c\Delta x})$ is given by

$$\big(d_{c\Delta x}\big)^2 u(x) = \frac{1}{(2\Delta x)^2}\big[u(x+2\Delta x) + u(x-2\Delta x) - 2u(x)\big]. \tag{7.15}$$

On the other hand, another approximation for the second derivative of a function can be obtained directly from the Taylor expansion formulas in (7.13)-(7.14). Indeed, add Eqs. (7.13)-(7.14), that is,

$$u(x+\Delta x) + u(x-\Delta x) = 2u(x) + \frac{d^2u}{dx^2}(x)\,(\Delta x)^2 + O\big((\Delta x)^3\big).$$

This equation can be rewritten as

$$\frac{d^2u}{dx^2}(x) = \frac{u(x+\Delta x) + u(x-\Delta x) - 2u(x)}{(\Delta x)^2} + O\big(\Delta x\big).$$

This equation suggests to introduce a second-order centered difference quotient $(d^2)_{c\Delta x}$ as

$$(d^2)_{c\Delta x}u(x) = \frac{1}{(\Delta x)^2}\big[u(x+\Delta x) + u(x-\Delta x) - 2u(x)\big]. \tag{7.16}$$

Using this notation, the equation above is given by

$$\frac{d^2u}{dx^2}(x) = (d^2)_{c\Delta x}u(x) + O\big(\Delta x\big).$$

Therefore, both $\big(d_{c\Delta x}\big)^2$ and $(d^2)_{c\Delta x}$ are approximations of the second derivative of a function. However, they are not the same approximation, since comparing Eqs. (7.15) and (7.16) it is not difficult to see that

$$\big(d_{c(\Delta x)/2}\big)^2 = (d^2)_{c\Delta x}.$$

We conclude that there are many different ways to approximate second derivatives by difference quotients, many more ways than those to approximate first order derivatives. In this Section we use the difference quotient in Eq. (7.16), and we use the simplified notation given by $d_c^2 = (d^2)_{c\Delta x}$.

7.3.3. **Method of finite differences.** We now describe the finite difference method using two examples. In the first example we find an approximate solution for the boundary value problem in Eqs. (7.7)-(7.8). In the second example we find an approximate solution for the initial-boundary value problem in Eqs. (7.9)-(7.11).

EXAMPLE 7.3.1: Consider the boundary value problem for the ordinary differential equation given in Eqs. (7.7)-(7.8).

(a) Divide the interval $[0,1]$ into $n > 1$ equal intervals and use the finite difference method to find a vector $\mathsf{u} = [u_i] \in \mathbb{R}^{n+1}$ that approximates the function $u : [0,1] \to \mathbb{R}$ solution of that boundary value problem. Use centered difference quotients to approximate the first and second derivatives of the unknown function u.

(b) Find the explicit form of the linear system in the case $n = 6$.

(c) Find the degree n polynomial p_n that interpolates the approximate solution vector $\mathsf{u} = [u_i] \in \mathbb{R}^{n+1}$, which includes the boundary points.

SOLUTION:

Part (a): Fix a positive integer $n \in \mathbb{N}$, define the grid step size $h = 1/n$, and introduce the uniform grid

$$\{x_i\}, \qquad x_i = ih, \qquad i = 0, 1, \cdots, n, \qquad \text{on } [0,1].$$

(See Fig. 50.) Introduce the numbers $f_i = f(x_i)$. Finally, denote $u_i = u(x_i)$. The numbers x_i and f_i are known from the problem, and so are the $u_0 = u(0) = 0$ and $u_n = u(1) = 0$,

while the u_i for $i = 1, \cdots, n-1$ are the unknowns. We now use the original differential equation to construct an $(n-1) \times (n-1)$ linear system for these unknowns u_i.

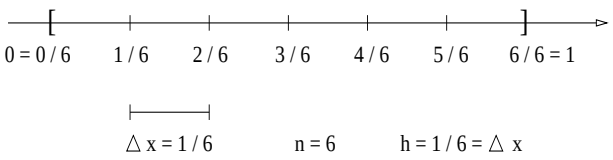

FIGURE 50. A uniform grid for $n = 6$ on the domain $[0, 1]$.

We use centered difference quotient given in Eqs. (7.12) and (7.16) to approximate the first and second derivatives of the function u, respectively. We choose $\Delta x = h$, and we denote the difference quotients evaluated at grid points x_i as follows,

$$d_c u(x_i) = d_c u_i, \qquad d_c^2 u(x_i) = d_c^2 u_i.$$

Therefore, we obtain the following formulas for the difference quotients,

$$d_c u_i = \frac{u_{i+1} - u_{i-1}}{2h}, \qquad d_c^2 u_i = \frac{u_{i+1} + u_{i-1} - 2u_i}{h^2}. \tag{7.17}$$

Now we state the approximate problem we will solve: Given the constants $\{f_i\}_{i=1}^{n-1}$, find the vector $\mathsf{u} = [u_i] \in \mathbb{R}^{n+1}$ solution of the $(n-1) \times (n-1)$ linear system and boundary conditions, respectively,

$$d_c^2 u_i + d_c u_i = f_i, \quad i = 1, \cdots, n-1, \tag{7.18}$$

$$u_0 = u_n = 0. \tag{7.19}$$

Eq. (7.18) is indeed a linear system for $\mathsf{u} = [u_i]$, since it is equivalent to the system

$$(2+h)u_{i+1} - 4u_i + (2-h)u_{i-1} = 2h^2 f_i, \qquad i = 1, \cdots, n-1.$$

When the boundary conditions in Eq. (7.19) are introduced in the equation above, we obtain an $(n-1) \times (n-1)$ linear system for the unknowns u_i, where $i = 1, \cdots, (n-1)$.

Part (b): In the case $n = 6$, we have $h = 1/6$, so we denote $a = 2 - 1/6$, $b = 2 + 1/6$ and $c = 2/36$. Also recall the boundary conditions $u_0 = u_6 = 0$. Then, the system above and its augmented matrix are given by, respectively,

$$\begin{aligned} -4u_1 + bu_2 &= c\,f_1, \\ au_1 - 4u_2 + bu_3 &= c\,f_2, \\ au_2 - 4u_3 + bu_4 &= c\,f_3, \\ au_3 - 4u_4 + bu_5 &= c\,f_4, \\ au_4 - 4u_5 &= c\,f_5, \end{aligned} \quad \Leftrightarrow \quad \left[\begin{array}{ccccc|c} -4 & b & 0 & 0 & 0 & c\,f_1 \\ a & -4 & b & 0 & 0 & c\,f_2 \\ 0 & a & -4 & b & 0 & c\,f_3 \\ 0 & 0 & a & -4 & b & c\,f_4 \\ 0 & 0 & 0 & a & -4 & c\,f_5 \end{array}\right].$$

We then conclude that the solution u of the boundary value problem in Eq. (7.7) can be approximated by the solution $\mathsf{u} = [u_i] \in \mathbb{R}^7$ of the 5×5 linear system above plus the two boundary conditions. The same type of approximate solution can be found for all $n \in \mathbb{N}$.

Part (c): The output of the finite difference method is a vector $\mathsf{u} = [u_i] \in \mathbb{R}^{(n+1)}$. An approximate solution to the ordinary differential equation in Eqs. (7.7) can be constructed from the vector u in many different ways. One way is polynomial interpolation, that is,

to construct a polynomial of degree n whose graph contains all the points (x_i, u_i). Such polynomial is given by

$$p_n(x) = \sum_{i=0}^{n} u_i\, q_i(x), \qquad q_i(x) = \prod_{j\neq i} \frac{(x - x_j)}{(x_i - x_j)}.$$

It can be verified that the degree n polynomials q_i when evaluated at grid points satisfies

$$q_i(x_j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Therefore, the polynomial p_n has degree n and satisfies that $p_n(x_i) = u_i$. This polynomial function p_n approximates the solution u of the boundary value problem in Eqs. (7.7)-(7.8). ◁

EXAMPLE 7.3.2: Consider the boundary value problem for the partial differential equation given in Eqs. (7.9)-(7.11). Use the finite difference method to find an approximate solution of the function $u : D \to \mathbb{R}$ solution of that initial-boundary value problem.

(a) Use centered difference quotients to approximate the spatial derivatives and forward difference quotients to approximate time derivatives of the unknown function u.
(b) Repeat the calculations in part (a) now using backward difference quotients for the time derivatives of the unknown function u.

SOLUTION:

Part (a): Introduce a grid in the domain $D = [0, \pi] \times [0, T]$ as follows: Fix the positive integers n_x, $n_t \in \mathbb{N}$, define the step sizes $h_x = \pi/n_x$ and $h_t = T/n_t$, and then introduce the uniform grids

$$\begin{aligned} &\{x_i\}, && x_i = ih_x, && i = 0, 1, \cdots, n_x, && \text{on } [0, \pi] \\ &\{t_j\}, && t_j = jh_t, && j = 0, 1, \cdots, n_t, && \text{on } [0, T]. \end{aligned}$$

A point of the form $(x_i, t_j) \in D$ is called a grid point. (See Fig. 51.)

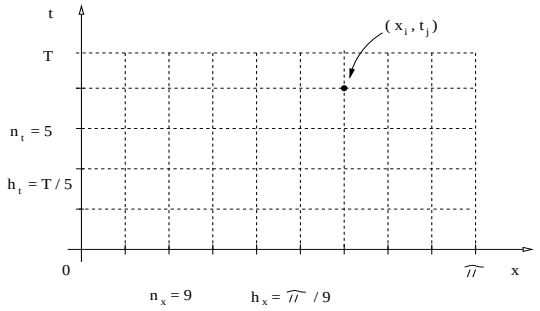

FIGURE 51. A rectangular grid with $n_x = 9$ and $n_t = 5$ on the domain $D = [0, \pi] \times [0, T]$.

We will compute the approximate solution values at grid points, and we use the notation $u_{i,j} = u(x_i, t_j)$ and $f_{i,j} = f(x_i, t_j)$ to denote unknown and source function values at grid points. We also denote $g_i = g(x_i)$ for the initial data function values at grid points. In this notation the boundary conditions in Eq. (7.10) have the form $u_{0,j} = 0$ and $u_{n_x,j} = 0$. We

finally introduce the forward difference quotient in time d_{+t} and the second order centered difference quotient in space d^2_{cx} as follows,

$$d_{+t}u_{i,j} = \frac{u_{i,(j+1)} - u_{i,j}}{h_t}, \qquad d^2_{cx}u_{i,j} = \frac{u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}}{h_x^2}.$$

Now we state the approximate problem we will solve: Given the constants $f_{i,j}$ and g_i, for $i = 0, 1, \cdots, n_x$ and $j = 0, 1, \cdots, n_y$, find the constants $u_{i,j}$ solution of the linear equations and boundary conditions, respectively,

$$d_{+t}u_{i,j} - \kappa d^2_{cx}u_{i,j} = f_{i,j}, \tag{7.20}$$

$$u_{0,j} = u_{n_x,j} = 0, \tag{7.21}$$

$$u_{i,0} = g_i. \tag{7.22}$$

This system is simpler to solve than it looks at first sight. Let us rewrite it in as follows,

$$\frac{1}{h_t}\big(u_{i,(j+1)} - u_{i,j}\big) - \frac{\kappa}{h_x^2}\left(u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}\right) = f_{ij},$$

which is equivalent to the equation

$$u_{i,(j+1)} = r\big(u_{(i+1),j} + u_{(i-1),j}\big) + (1 - 2r)\, u_{i,j} + h_t f_{i,j}, \tag{7.23}$$

where $r = \kappa h_t/h^2_{cx}$. This last equation says that the solution at the time t_{j+1} can be computed if the solution at the previous time t_j is known. Since the solution at the initial time $t_0 = 0$ is known an equal to g_i, and since solution at the boundary of the domain $u_{0,j}$ and $u_{n_x,j}$ is known from the boundary conditions, then the solution $u_{i,j}$ can be computed from Eq. (7.23) time step after time step. For this reason the linear system in Eqs. (7.20)-(7.22) above is an example of an *explicit method*.

Part (b): We now repeat the calculations in part (a) using a backward difference quotient in time. We introduce the notation d_{-t} for the backward difference quotient in time, and we keep the notation d^2_{cx} for the second order centered difference quotient in space,

$$d_{-t}u_{i,j} = \frac{u_{i,j} - u_{i,(j-1)}}{h_t}, \qquad d^2_{cx}u_{i,j} = \frac{u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}}{h_x^2}.$$

Now we state the approximate problem we will solve: Given the constants $f_{i,j}$ and g_i, for $i = 0, 1, \cdots, n_x$ and $j = 0, 1, \cdots, n_y$, find the constants $u_{i,j}$ solution of the linear equations and boundary conditions, respectively,

$$d_{-t}u_{i,j} - \kappa d^2_{cx}u_{i,j} = f_{i,j}, \tag{7.24}$$

$$u_{0,j} = u_{n_x,j} = 0, \tag{7.25}$$

$$u_{i,0} = g_i. \tag{7.26}$$

This system is simpler to solve than it looks at first sight. Let us rewrite it in as follows,

$$\frac{1}{h_t}\big(u_{i,j} - u_{i,(j-1)}\big) - \frac{\kappa}{h_x^2}\left(u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}\right) = f_{ij},$$

which is equivalent to the equation

$$(+2r)u_{i,j} - r\big(u_{(i+1),j} + u_{(i-1),j}\big) = u_{i,(j-1)} + h_t f_{i,j}, \tag{7.27}$$

where $r = \kappa h_t/h^2_{cx}$, as above. This last equation says that the solution at the time t_{j+1} can be computed if the solution at the previous time t_j is known. However, in this case we need to solve an $(n_x - 1) \times (n_x - 1)$ linear system at each time step. Such system is similar to the one that appeared in Example 7.3.1. Since the solution at the initial time $t_0 = 0$ is known an equal to g_i, and since solution at the boundary of the domain $u_{0,j}$ and $u_{n_x,j}$ is known from the boundary conditions, then the solution $u_{i,j}$ can be computed from Eq. (7.27) time step

after time step. We emphasize that the solution is computed by solving an $(n_x-1)\times(n_x-1)$ system at every time step. For this reason the linear system in Eqs. (7.24)-(7.26) above is an example of an *implicit method*. ◁

What we have seen in this Section is just the first part of the story. We have seen how we can use linear algebra to obtain approximate solutions of few problems involving differential equations. The second part is to study how the solutions of the approximate problems approaches the solution of the original problem as the grid step size approaches zero. Consider the approximate solution $\mathsf{u} \in \mathbb{R}^{n+1}$ found in Example 7.3.1. Does the interpolation polynomial p_n constructed with the components of u approximate the function $u : [0,1] \to \mathbb{R}$ solution to the boundary value problem in Eq. (7.7) in the limit $n \to \infty$? A similar question can be asked for the solutions $\{u_{i,j}\}$ obtained in parts (b) and (c) in Example 7.3.2. We will study the answers to these questions in the following Chapters.

One last remark is the following. Comparing parts (a) and (b) in Example 7.3.2 we see that explicit methods are simpler to solve than implicit methods. A matrix must be inverted to solve an implicit method, while this is not needed to solve an explicit method. So, why are implicit methods studied at all? The reason is that in the limit $n \to \infty$ the approximate solutions of explicit methods do not approximate the solution of the original differential equation as good as a solution of an implicit method. Moreover, the solution of an explicit method may not converge at all, while the solutions of implicit methods always converges.

Further reading. See Section 1.4 in Meyer's book [3] for a detailed discussion on discretizations of two-point boundary values problems.

7.3.4. **Exercises.**

7.3.1.- Consider the boundary value problem for the function u given by

$$\frac{d^2u}{dx^2}(x) = 25\,x,$$
$$u(0) = 0, \quad u(1) = 0,$$
$$x \in [0,1].$$

Divide the interval $[0,1]$ into five equal subintervals and use the finite difference method to find an approximate solution vector $\mathsf{u} = [u_i]$ to the boundary value problem above, where $i = 0, \cdots, 5$. Use **centered** difference quotients given in Eq. (7.17) to approximate the derivatives of function u.

7.3.2.- Given an infinite differentiable function $u : \mathbb{R} \to \mathbb{R}$. apply twice the forward difference quotient d_+ to show that the second order forward difference quotient has the form

$$d_+^2 u(x) = \frac{u(x+2\Delta x) - 2u(x+\Delta x) + u(x)}{(\Delta x)^2}.$$

7.3.3.- Given an infinite differentiable function $u : \mathbb{R} \to \mathbb{R}$. apply twice the backward difference quotient d_+ to show that the second order backward difference quotient has the form

$$d_-^2 u(x) = \frac{u(x) - 2u(x-\Delta x) + u(x-2\Delta x)}{(\Delta x)^2}.$$

7.3.4.- Consider the boundary value problem given in the Exercise **7.3.1**. Divide again the interval $[0,1]$ into five equal subintervals and find the 4×4 linear system that approximates the original problem using **forward** difference quotients to approximate derivatives of function u. You do not need to solve the linear system.

7.3.5.- Consider the boundary value problem given Problem **7.3.1**. Divide again the interval $[0,1]$ into five equal subintervals and find the 4×4 linear system that approximates the original problem using **backward** difference quotients to approximate derivatives of function u. You do not need to solve the linear system.

7.3.6.- Consider the boundary value problem for the function u given by

$$\frac{d^2u}{dx^2}(x) + 2\frac{du}{dx}(x) = 25\,x,$$
$$u(0) = 0, \quad u(1) = 0,$$
$$x \in [0,1].$$

Divide the interval $[0,1]$ into five equal subintervals and use the finite difference method to find an algebraic linear system that approximates the original boundary value problem above. Use centered difference quotients given in Eq. (7.17) to approximate derivatives of function u. You do not need to solve the linear system.

7.4. Finite element method

The finite element method permits the computation of approximate solutions to differential equations. Boundary value problems involving a differential equations are first transformed into integral equations. The boundary conditions are included into the integral equation by performing integration by parts. The original problem is transformed into inverting a bilinear form on a vector space. The approximate problem is obtained when the integral equation is solved not on the whole vector space but on a finite dimensional subspace. By a careful choice of the subspace, the calculations needed to obtain the approximate solution can be simplified. In this Section we study the same differential equations we have seen in Sect. 7.3 when we described the finite difference methods. Finite difference methods are different form finite element methods. The former method approximates derivatives by difference quotients in order to obtain the approximate problem involving an algebraic linear system. The latter method produces an algebraic linear system by restricting an integral version of the original differential equation onto a finite dimensional subspace of the original vector space where the integral equation is defined.

7.4.1. **Differential equations.** We now recall the boundary value problem we are interested to study. This is the first problem we studied in Sect. 7.3, that is, to find an approximate solution to a boundary value problem for an ordinary differential equation: Given an infinitely differentiable function $f : [0,1] \to \mathbb{R}$, find a function $u : [0,1] \to \mathbb{R}$ solution of the boundary value problem

$$\frac{d^2u}{dx^2}(x) + \frac{du}{dx}(x) = f(x), \tag{7.28}$$

$$u(0) = u(1) = 0. \tag{7.29}$$

Recall that finding the function u involves doing two integrations that introduce two integration constants. These constants are determined by the two conditions on $x = 0$ and $x = 1$ above, called boundary conditions, since they are conditions on the boundaries of the interval $[0,1]$. Because of this extra condition on the ordinary differential equation this problem is called a boundary value problem.

This problem can be expressed using linear transformations on vector spaces. Consider the inner product spaces $\big(V, \langle\,,\rangle\big)$ and $\big(W, \langle\,,\rangle\big)$, where $V = C_0^\infty([0,1],\mathbb{R})$ is the space of infinitely many differentiable functions that vanish at $x = 0$ and $x = 1$, $W = C^\infty([0,1],\mathbb{R})$ is the space of infinitely many differentiable functions, while the inner product is defined as

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \int_0^1 \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx.$$

Introduce the linear transformation $L : V \to W$ defined by

$$L(\boldsymbol{v}) = \frac{d^2\boldsymbol{v}}{dx^2} + \frac{d\boldsymbol{v}}{dx}$$

Then, the boundary value problem in Eqs. (7.28)-(7.29) can be expressed in the following way: Given a vector $\boldsymbol{f} \in W$, find a vector $\boldsymbol{u} \in V$ solution of the equation

$$L(\boldsymbol{u}) = \boldsymbol{f}. \tag{7.30}$$

Notice that the boundary conditions in Eq. (7.29) have been included into the definition of the vector space V. The problem defined by Eqs. (7.28)-(7.29), which is equivalently defined by Eq. (7.30), is called the **strong formulation** of the problem.

So, we have expressed a boundary value problem for a differential equation in terms of linear transformations between appropriate infinite dimensional vector spaces. The next step is to use the inner product defined on V and W to transform the differential equation

in (7.30) into an integro-differential equation, which will be called the weak formulation of the problem. This idea is summarized below.

7.4.2. **The Galerkin method.** The Galerkin method refers to a collection of ideas to transform a problem involving a linear transformation between infinite dimensional inner product spaces into a problem involving a matrix as a function between finite dimensional subspaces. We describe in this Section the original idea, introduced by Boris Galerkin around 1915. Galerkin worked only with partial differential equations, but we now know that his idea works in the more general context of a linear transformation between infinite dimensional inner product spaces. For this reason we describe Galerkin's idea in this more general context. The Galerkin method is to transform the strong formulation of the problem in (7.30) into what is called the weak formulation of the problem. This transformation is done using the inner product defined on the infinite dimensional vector spaces. Before we describe this transformation we need few definitions.

7.4.3. **Finite element method.**

7.4.4. **Exercises.**

7.4.1.- .

7.4.2.- .

Chapter 8. Normed spaces

8.1. The p-norm

An inner product in a vector space always determines an inner product norm, which satisfies the properties (a)-(c) in Theorem 6.2.5. However, the inner product norm is not the only function $V \to \mathbb{R}$ satisfying these properties.

Definition 8.1.1. *A* **norm** *on a vector space V over the field $\mathbb{F}$ is any function function $\| \ \| : V \to \mathbb{R}$ satisfying the following properties,*

(a) **(Positive definiteness)** *For all $\boldsymbol{x} \in V$ holds $\|\boldsymbol{x}\| \geqslant 0$, and $\|\boldsymbol{x}\| = 0$ iff $\boldsymbol{x} = 0$;*
(b) **(Scaling)** *For all $\boldsymbol{x} \in V$ and all $a \in \mathbb{F}$ holds $\|a\boldsymbol{x}\| = |a| \, \|\boldsymbol{x}\|$;*
(c) **(Triangle inequality)** *For all $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.*

A **normed space** *is a pair $\big(V, \| \ \|\big)$ of a vector space with a norm.*

The inner product norm, $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ for all $\boldsymbol{x} \in V$, defined in an inner product space $\big(V, \langle \, , \rangle\big)$, is thus an obvious example of a norm, since it satisfies the properties (a)-(c) in Theorem 6.2.5, which are precisely the conditions given in Definition 8.1.1. It is important to notice that alternative norms exist on inner product spaces. Moreover, a norm can be introduced in a vector space without having an inner product structure. One particularly important example of the former case is given by the p-norms defined on $V = \mathbb{F}^n$.

Definition 8.1.2. *The* **p-norm** *on the vector space $\mathbb{F}^n$, with $1 \leqslant p \leqslant \infty$, is the function $\| \ \|_p : \mathbb{F}^n \to \mathbb{R}$ defined as follows,*

$$\|\mathsf{x}\|_p = \big(|x_1|^p + \cdots + |x_n|^p\big)^{1/p}, \qquad p \in [1, \infty),$$
$$\|\mathsf{x}\|_\infty = \max\big\{|x_1|, \cdots, |x_n|\big\}, \qquad (p = \infty),$$

with $\mathsf{x} = [x_i]$, for $i = 1, \cdots, n$, the vector components in the standard ordered basis of $\mathbb{F}^n$.

Since the dot product norm is given by $\|\mathsf{x}\| = \big(|x_1|^2 + \cdots + |x_n|^2\big)^{1/2}$, it is simple to see that $\| \ \| = \| \ \|_2$, that is, the case $p = 2$ coincides with the dot product norm on $\mathbb{F}^n$. The most commonly used norms, besides $p = 2$, are the cases $p = 1$ and $p = \infty$,

$$\|\mathsf{x}\|_1 = |x_1| + \cdots + |x_n|, \qquad \|\mathsf{x}\|_\infty = \max\big\{|x_1|, \cdots, |x_n|\big\}.$$

Theorem 8.1.3. *For each value of $p \in [1, \infty]$ the p-norm function $\| \ \|_p : \mathbb{F}^n \to \mathbb{R}$ introduced in Definition 8.1.2 is a norm on $\mathbb{F}^n$.*

The Theorem above states that the function $\| \ \|_p$ satisfies the properties (a)-(c) in Definition 8.1.1. Therefore, for each value $p \in [1, \infty]$ the space $\big(\mathbb{F}^n, \| \ \|_p\big)$ is a normed space. We will see at the end of this Section that in the case $p \in [1, \infty]$ and $p \neq 2$ these norms are not inner product norms. In other words, for these values of p there is **no** inner product $\langle \, , \rangle_p$ defined on $\mathbb{F}^n$ such that $\| \ \|_p = \sqrt{\langle \, , \rangle_p}$.

In order to prove that the p-norm is indeed a norm, we first need to establish the following generalization of the Cauchy-Schwarz inequality, called Hölder inequality.

Theorem 8.1.4. (Hölder inequality) *For all vectors x, $\mathsf{y} \in \mathbb{F}^n$ and $p \in [1, \infty)$ holds that*

$$|\mathsf{x}^* \mathsf{y}| \leqslant \|\mathsf{x}\|_p \, \|\mathsf{y}\|_q, \qquad \text{with} \qquad \frac{1}{p} + \frac{1}{q} = 1. \tag{8.1}$$

Proof of Theorem 8.1.4: We first show that for all real numbers $a \geqslant 0$, $b > 0$ holds

$$a^\lambda b^{1-\lambda} \leqslant (1 - \lambda) \, b + \lambda \, a \qquad \lambda \in [0, 1]. \tag{8.2}$$

This inequality can be shown using the auxiliary function

$$f(t) = (1-\lambda) + \lambda\, t - t^{\lambda}.$$

Its derivative is $f'(t) = \lambda[1 - t^{(\lambda-1)}]$, so the function f satisfies

$$f(1) = 0, \qquad f'(t) > 0 \quad \text{for} \quad t > 1, \qquad f'(t) < 0 \quad \text{for} \quad 0 \leqslant t < 1.$$

We conclude that $f(t) \geqslant 0$ for $t \in [0, \infty)$. Given two real numbers $a \geqslant 0$, $b > 0$, we have proven that $f(a/b) \geqslant 0$, that is,

$$\frac{a^{\lambda}}{b^{\lambda}} \leqslant (1-\lambda) + \lambda \frac{a}{b} \quad \Leftrightarrow \quad a^{\lambda} b^{1-\lambda} \leqslant (1-\lambda)\, b + \lambda\, a.$$

Having established Eq. (8.2), we use it to prove Hölder inequality. Let $\mathsf{x} = [x_i]$, $\mathsf{y} = [y_j] \in \mathbb{F}^n$ be arbitrary vectors, where $i, j = 1, \cdots, n$, and introduce the rescaled vectors

$$\hat{\mathsf{x}} = \frac{\mathsf{x}}{\|\mathsf{x}\|_p}, \qquad \hat{\mathsf{y}} = \frac{\mathsf{y}}{\|\mathsf{y}\|_q}, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

These rescaled vectors satisfy that $\|\hat{\mathsf{x}}\|_p = 1$ and $\|\hat{\mathsf{y}}\|_q = 1$. Denoting $\hat{\mathsf{x}} = [\hat{x}_i]$ and $\hat{\mathsf{y}} = [\hat{y}_i]$, use the inequality in Eq. (8.2) in the case that $a = \hat{x}_i$, $b = \hat{y}_i$ and $\lambda = 1/p$, as follows,

$$\begin{aligned} |\overline{\hat{x}}_i \hat{y}_i| &= \Big(\frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}\Big)^{\frac{1}{p}} \Big(\frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}\Big)^{1-\frac{1}{p}} \\ &\leqslant \Big(1 - \frac{1}{p}\Big)\Big(\frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}\Big) + \frac{1}{p}\Big(\frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}\Big) \\ &\leqslant \frac{1}{q}\,|\hat{y}_i|^q + \frac{1}{p}\,|\hat{x}_i|^p. \end{aligned}$$

Adding up over all components,

$$|\hat{\mathsf{x}}^* \hat{\mathsf{y}}| \leqslant \sum_{i=1}^n |\overline{\hat{x}}_i \hat{y}_i| \leqslant \frac{1}{q}\,|\hat{y}_i|^q + \frac{1}{p}\,|\hat{x}_i|^p = \frac{1}{q}\,\|\hat{\mathsf{y}}\|_q^q + \frac{1}{p}\,\|\hat{\mathsf{x}}\|_p^p = \frac{1}{q} + \frac{1}{p} = 1.$$

Therefore, $|\hat{\mathsf{x}}^* \hat{\mathsf{y}}| \leqslant 1$, which is equivalent to

$$\Big|\frac{\mathsf{x}^*}{\|\mathsf{x}\|_p}\,\frac{\mathsf{y}}{\|\mathsf{y}\|_q}\Big| \leqslant 1 \quad \Leftrightarrow \quad |\mathsf{x}^* \mathsf{y}| \leqslant \|\mathsf{x}\|_p\, \|\mathsf{y}\|_q.$$

This establishes the Theorem. □

The Hölder inequality plays an important role to show that the p-norms satisfy the triangle inequality. We saw the same situation when the Cauchy-Schwarz inequality played a crucial role to prove that the inner product norm satisfied the triangle inequality.

Proof of Theorem 8.1.3: We show the proof for $p \in [1, \infty)$. The case $p = \infty$ is left as an exercise. So, we assume that $p \in [1, \infty)$, we introduce q by the equation $\frac{1}{p} + \frac{1}{q} = 1$. In order to show that the p-norm is a norm we need to show that the p-norm satisfies the properties (a)-(c) in Definition 8.1.1. The first two properties a simple to prove. The p-norm is positive, since for all $\mathsf{x} \in \mathbb{F}^n$ holds

$$\|\mathsf{x}\|_p = \Big(\sum_{i=1}^n |x_i|^p\Big)^{\frac{1}{p}} \geqslant 0, \quad \text{and} \quad \Big(\sum_{i=1}^n |x_i|^p\Big)^{\frac{1}{p}} = 0 \quad \Leftrightarrow \quad |x_i| = 0 \quad \Leftrightarrow \quad \mathsf{x} = \mathsf{0}.$$

The p-norm satisfies the scaling property, since for all $\mathsf{x} \in \mathbb{F}^n$ and all $a \in \mathbb{F}$ holds

$$\|a\mathsf{x}\|_p = \Big(\sum_{i=1}^n |a x_i|^p\Big)^{\frac{1}{p}} = \Big(\sum_{i=1}^n |a|^p\, |x_i|^p\Big)^{\frac{1}{p}} = \Big(|a|^p \sum_{i=1}^n |x_i|^p\Big)^{\frac{1}{p}} = |a|\, \|\mathsf{x}\|_p.$$

The difficult part is to establish that the p-norm satisfies the triangle inequality. We start proving the following statement: For all real numbers a, b holds

$$|a+b|^p \leqslant |a|\,|a+b|^{\frac{p}{q}} + |b|\,|a+b|^{\frac{p}{q}}. \tag{8.3}$$

This is indeed the case, since

$$|a+b|^p = |a+b|\,|a+b|^{p-1}, \quad \text{and} \quad \frac{1}{q} = 1 - \frac{1}{p} = \frac{p-1}{p} \quad \Leftrightarrow \quad \frac{p}{q} = p-1,$$

so we conclude that

$$|a+b| = |a+b|\,|a+b|^{\frac{p}{q}} \leqslant (|a|+|b|)\,|a+b|^{\frac{p}{q}},$$

which is the inequality in Eq. (8.3). This inequality will be used in the following calculation. Given arbitrary vectors $\mathsf{x} = [x_i]$, $\mathsf{y} = [y_i] \in \mathbb{F}^n$, for $i = 1, \cdots, n$, the following inequalities hold

$$\|\mathsf{x}+\mathsf{y}\|_p^p = \sum_{i=1}^n |x_i + y_i|^p \leqslant \sum_{i=1}^n \Big(|x_i|\,|x_i+y_i|^{\frac{p}{q}} + |y_i|\,|x_i+y_i|^{\frac{p}{q}} \Big), \tag{8.4}$$

where Eq. (8.3) was used to obtain the last inequality. Now the Hölder inequality implies both

$$\sum_{i=1}^n |x_i|\,|x_i+y_i|^{\frac{p}{q}} \leqslant \Big(\sum_{i=1}^n |x_i|^p \Big)^{\frac{1}{p}} \Big(\sum_{i=1}^n |x_i+y_i|^p \Big)^{\frac{1}{q}},$$
$$\sum_{i=1}^n |y_i|\,|x_i+y_i|^{\frac{p}{q}} \leqslant \Big(\sum_{i=1}^n |y_i|^p \Big)^{\frac{1}{p}} \Big(\sum_{i=1}^n |x_i+y_i|^p \Big)^{\frac{1}{q}}.$$

Inserting these expressions in Eq. (8.4) we obtain

$$\|\mathsf{x}+\mathsf{y}\|_p^p \leqslant \|\mathsf{x}\|_p \big(\|\mathsf{x}+\mathsf{y}\|_p \big)^{\frac{p}{q}} + \|\mathsf{y}\|_p \big(\|\mathsf{x}+\mathsf{y}\|_p \big)^{\frac{p}{q}}.$$

Recalling that $\frac{p}{q} = p-1$, we obtain the inequality

$$\|\mathsf{x}+\mathsf{y}\|_p^p \leqslant \big(\|\mathsf{x}\|_p + \|\mathsf{y}\|_p \big) \|\mathsf{x}+\mathsf{y}\|_p^{(p-1)} \quad \Leftrightarrow \quad \|\mathsf{x}+\mathsf{y}\|_p \leqslant \|\mathsf{x}\|_p + \|\mathsf{y}\|_p.$$

This establishes the Theorem for $p \in [1, \infty)$. □

EXAMPLE 8.1.1: Find the length of $\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ in the norms $\|\ \|_1$, $\|\ \|_2$ and $\|\ \|_\infty$.

SOLUTION: A simple calculations shows,

$$\|\mathsf{x}\|_1 = |1| + |2| + |-3| = 6,$$
$$\|\mathsf{x}\|_2 = \big[1^2 + 2^2 + (-3)^2\big]^{1/2} = \sqrt{14} = 3.74,$$
$$\|\mathsf{x}\|_\infty = \max\big\{|1|, |2|, |-3|\big\} = 3.$$

Notice that for the vector x above holds

$$\|\mathsf{x}\|_\infty \leqslant \|\mathsf{x}\|_2 \leqslant \|\mathsf{x}\|_1. \tag{8.5}$$

One can prove that the inequality in Eq. (8.5) holds for all $\mathsf{x} \in \mathbb{F}^n$. ◁

EXAMPLE 8.1.2: Sketch on $\mathbb{R}^2$ the set of vectors $B_p = \big\{\mathsf{x} \in \mathbb{R}^2 \,:\, \|\mathsf{x}\|_p = 1\big\}$ for the cases $p = 1$, $p = 2$, and $p = \infty$.

SOLUTION: Recall we use the standard basis to express $\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. We start with the set B_2, which is the circle of radius one in Fig. 52, that is,

$$(x_1)^2 + (x_2)^2 = 1.$$

The set B_1 is the square of side one given by

$$|x_1| + |x_2| = 1.$$

The sides are given by the lines $\pm x_1 \pm x_2 = 1$. See Fig. 52. The set B_∞ is the square of side two given by

$$\max\{|x_1|, |x_2|\} = 1.$$

The sides are given by the lines $x_1 = \pm 1$ and $x_2 = \pm 1$. See Fig. 52.

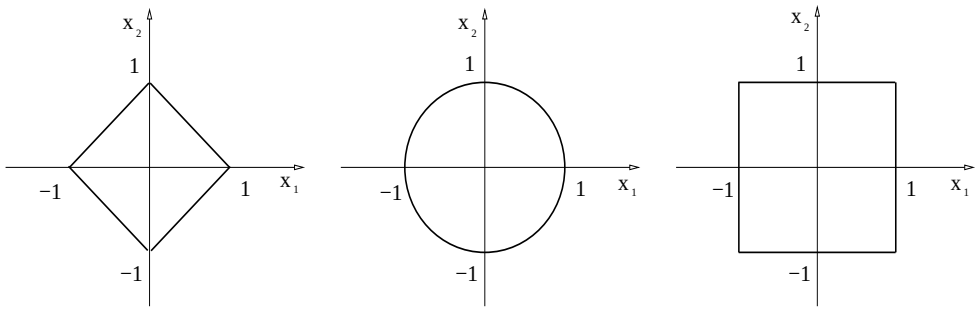

FIGURE 52. Unit sets in $\mathbb{R}^2$ for the p-norms, with $p = 1, 2, \infty$, respectively.

◁

EXAMPLE 8.1.3: Show that for every $\mathsf{x} \in \mathbb{R}^2$ holds

$$\|\mathsf{x}\|_\infty \leqslant \|\mathsf{x}\|_2 \leqslant \|\mathsf{x}\|_1.$$

SOLUTION: Introducing the unit disks $D_p = \{\mathsf{x} \in \mathbb{R}^2 : \|\mathsf{x}\|_p \leqslant 1\}$, with $p = 1, 2, \infty$, then Fig. 53 shows that $D_1 \subset D_2 \subset D_\infty$. Let us choose $\mathsf{y} \in \mathbb{R}^2$ such that $\|\mathsf{y}\|_2 = 1$, that is, a vector on the circle, for example the vector given in second picture in Fig. 53. Since this vector is outside the disk D_1, that implies $\|\mathsf{y}\|_1 \geqslant 1$, and since this vector is inside the disk D_∞, that implies $\|\mathsf{y}\|_\infty \leqslant 1$. The three conditions together say

$$\|\mathsf{y}\|_\infty \leqslant \|\mathsf{y}\|_2 \leqslant \|\mathsf{y}\|_1.$$

The equal signs correspond to the cases where y is a horizontal or a vertical vector. Since any vector $\mathsf{x} \in \mathbb{R}^2$ is a scaling of an appropriate vector y on the border of D_2, that is, $\mathsf{x} = c\,\mathsf{y}$, with $0 \leqslant c \in \mathbb{R}$, then, multiplying the inequality above by c we obtain

$$\|\mathsf{x}\|_\infty \leqslant \|\mathsf{x}\|_2 \leqslant \|\mathsf{x}\|_1, \qquad \forall\, \mathsf{x} \in \mathbb{R}^2.$$

◁

The p-norms can be defined on infinite dimensional vector spaces like function spaces.

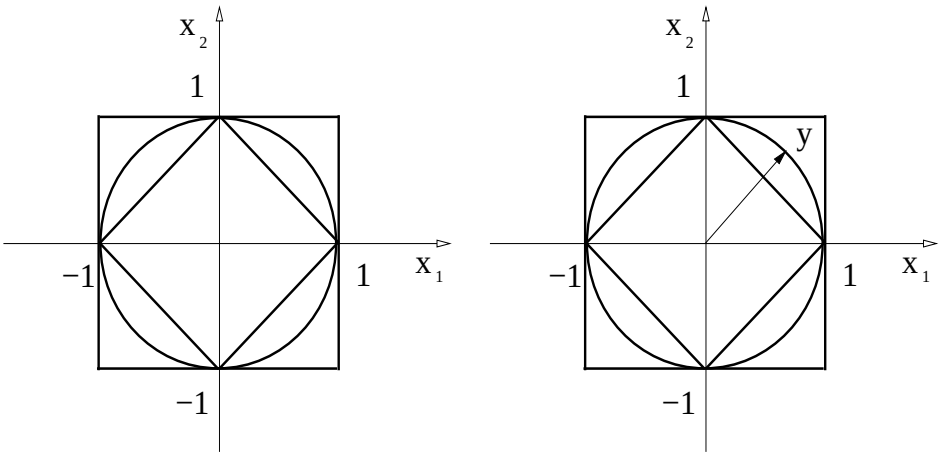

FIGURE 53. A comparison of the unit sets in $\mathbb{R}^2$ for the p-norms, with $p = 1, 2, \infty$.

Definition 8.1.5. *The* p**-norm** *on the vector space* $V = C^k([a,b], \mathbb{R})$*, with* $1 \leqslant p \leqslant \infty$*, is the function* $\| \ \|_p : V \to \mathbb{R}$ *defined as follows,*

$$\|\boldsymbol{f}\|_p = \Big(\int_a^b |\boldsymbol{f}(x)|^p \, dx \Big)^{1/p}, \qquad p \in [1, \infty),$$

$$\|\boldsymbol{f}\|_\infty = \max_{x \in [a,b]} |\boldsymbol{f}(x)|, \qquad (p = \infty).$$

One can show that the p-norms introduced in Definition 8.1.5 are indeed norms on the vector space $C^k([a,b], \mathbb{R})$. The proof of this statement follows the same ideas given in the proofs of Theorems 8.1.4 and 8.1.3 above, and we do not present it in these notes.

EXAMPLE 8.1.4: Consider the normed space $\big(C^0([-1,1], \mathbb{R}), \| \ \|_p\big)$ for any $p \in [1, \infty]$ and find the p-norm of the element $\boldsymbol{f}(x) = x$.

SOLUTION: In the case of $p \in [1, \infty)$ we obtain

$$\|\boldsymbol{f}\|_p^p = \int_{-1}^1 |x|^p \, dx = 2 \int_0^1 x^p \, dx = 2 \frac{x^{(p+1)}}{(p+1)} \Big|_0^1 = \frac{2}{p+1} \quad \Rightarrow \quad \|\boldsymbol{f}\|_p = \Big(\frac{2}{p+1} \Big)^{\frac{1}{p}}.$$

In the case of $p = \infty$ we obtain

$$\|\boldsymbol{f}\|_\infty = \max_{x \in [-1,1]} |x| \quad \Rightarrow \quad \|\boldsymbol{f}\|_\infty = 1.$$

Relations between the p-norms of a given vector $\boldsymbol{f}$ analogous to those relations found in Example 8.1.3 are not longer true. The volume of the integration region, the interval $[a,b]$, appears in any relation between p-norms of a fixed vector. We do not address such relations in these notes. ◁

8.1.1. **Not every norm is an inner product norm.** Since every inner product in a vector space determines a norm, the inner product norm, it is natural to ask whether the converse property holds: Does every norm in a vector space determine an inner product? The answer is no. Only those norms satisfying an extra condition, called the parallelogram identity, define an inner product.

Definition 8.1.6. *A norm* $\| \ \|$ *in a vector space* V *satisfies the* **polarization identity** *iff for all vectors* $\boldsymbol{x}$, $\boldsymbol{y} \in V$ *holds*

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right).$$

The polarization identity (also referred as parallelogram identity) is a well-known property of the dot product norm, which is geometrically described in Fig. 54 in the case of the vector space $\mathbb{R}^2$. It turns out that this property is crucial to determine whether a norm is an inner product norm for some inner product.

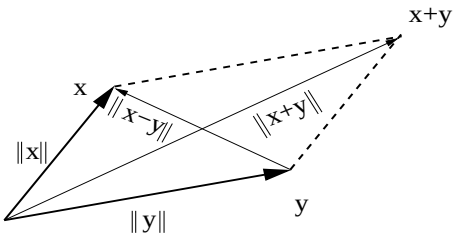

FIGURE 54. The polarization identity says that the sum of the squares of the diagonals in a parallelogram is twice the sum of the squares of the sides.

Theorem 8.1.7. *Given a normed space* $\big(V, \| \ \|\big)$, *the norm* $\| \ \|$ *is an inner product norm iff the norm* $\| \ \|$ *satisfies the polarization identity.*

It is not difficult to see that an inner product norm satisfies the polarization identity; see the first part in the proof below. It is rather involved to show the converse statement. If the norm $\| \ \|$ satisfies the polarization identity and V is a real vector space, then one shows that the function $\langle \, , \rangle : V \times V \to \mathbb{R}$ given by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\right)$$

is an inner product on V; in the case that V is a complex vector space, then one shows that the function $\langle \, , \rangle : V \times V \to \mathbb{C}$ given by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\right) + \frac{i}{4}\left(\|\boldsymbol{x}+i\boldsymbol{y}\|^2 - \|\boldsymbol{x}-i\boldsymbol{y}\|^2\right)$$

is an inner product on V.

Proof of Theorem 8.1.7:

($\Rightarrow$) Consider the inner product space $\big(V, \langle \, , \rangle\big)$ with inner product norm $\| \ \|$. For all vectors $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \langle (\boldsymbol{x}+\boldsymbol{y}), (\boldsymbol{x}+\boldsymbol{y}) \rangle = \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \|\boldsymbol{y}\|^2,$$
$$\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \langle (\boldsymbol{x}-\boldsymbol{y}), (\boldsymbol{x}-\boldsymbol{y}) \rangle = \|\boldsymbol{x}\|^2 - \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \|\boldsymbol{y}\|^2.$$

Adding both equations above up we obtain that

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right).$$

($\Leftarrow$) We only give the proof for real vector spaces. The proof for complex vector spaces is left as an exercise. Consider the normed space $\big(V, \| \ \|\big)$ and assume that V is a real vector space. In this case, introduce the function $\langle \, , \rangle : V \times V \to \mathbb{R}$ as follows,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\right).$$

Notice that this function satisfies $\langle \boldsymbol{x}, \boldsymbol{0}\rangle = \frac{1}{4}\left(\|\boldsymbol{x}\|^2 - \|\boldsymbol{x}\|^2\right) = 0$ for all $\boldsymbol{x} \in V$. We now show that this function $\langle\,,\rangle$ is an inner product on V. It is positive definite, since

$$\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \frac{1}{4}\|2\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$$

and the norm is positive definite, so we obtain that

$$\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geqslant 0, \quad \text{and} \quad \langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0 \quad \Leftrightarrow \quad \boldsymbol{x} = \boldsymbol{0}.$$

The function $\langle\,,\rangle$ is symmetric, since

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\right) = \frac{1}{4}\left(\|\boldsymbol{y}+\boldsymbol{x}\|^2 - \|\boldsymbol{y}-\boldsymbol{x}\|^2\right) = \langle \boldsymbol{y}, \boldsymbol{x}\rangle.$$

The difficult part is to show that the function $\langle\,,\rangle$ is linear in the second argument. Here is where we need the polarization identity. We start with the following expressions, which are obtained from the polarization identity,

$$\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2 + \|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|^2 = 2\,\|\boldsymbol{x}+\boldsymbol{y}\|^2 + 2\,\|\boldsymbol{z}\|^2,$$
$$\|\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{z}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}-\boldsymbol{z}\|^2 = 2\,\|\boldsymbol{x}-\boldsymbol{y}\|^2 + 2\,\|\boldsymbol{z}\|^2.$$

If we subtract the second equation from the first one, and ordering the terms conveniently, we obtain,

$$\Big[\|\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})\|^2 - \|\boldsymbol{x}-(\boldsymbol{y}+\boldsymbol{z})\|^2\Big] + \Big[\|\boldsymbol{x}+(\boldsymbol{y}-\boldsymbol{z})\|^2 - \|\boldsymbol{x}-(\boldsymbol{y}-\boldsymbol{z})\|^2\Big] = 2\Big[\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\Big]$$

which can be written in terms of the function $\langle\,,\rangle$ as follows,

$$\langle \boldsymbol{x}, (\boldsymbol{y}+\boldsymbol{z})\rangle + \langle \boldsymbol{x}, (\boldsymbol{y}-\boldsymbol{z})\rangle = 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle. \tag{8.6}$$

Several relations come from this equation. For the first relation, take $\boldsymbol{y} = \boldsymbol{z}$, and recall that $\langle \boldsymbol{x}, \boldsymbol{0}\rangle = 0$, then we obtain

$$\langle \boldsymbol{x}, 2\boldsymbol{y}\rangle = 2\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle. \tag{8.7}$$

The second relation derived from Eq. (8.6) is obtained renaming the vectors $\boldsymbol{y}+\boldsymbol{z} = \boldsymbol{u}$ and $\boldsymbol{y}-\boldsymbol{z} = \boldsymbol{v}$, that is,

$$\langle \boldsymbol{x}, \boldsymbol{u}\rangle + \langle \boldsymbol{x}, \boldsymbol{v}\rangle = 2\Big\langle \boldsymbol{x}, \frac{(\boldsymbol{u}+\boldsymbol{v})}{2}\Big\rangle = \langle \boldsymbol{x}, (\boldsymbol{u}+\boldsymbol{v})\rangle,$$

where the equation on the far right comes from Eq. (8.7). We have shown that for all $\boldsymbol{x}$, $\boldsymbol{u}$, $\boldsymbol{v} \in V$ holds

$$\langle \boldsymbol{x}, (\boldsymbol{u}+\boldsymbol{v})\rangle = \langle \boldsymbol{x}, \boldsymbol{u}\rangle + \langle \boldsymbol{x}, \boldsymbol{v}\rangle$$

which is a particular case of the linearity in the second argument property of an inner product. We only need to show that for all $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and all $a \in \mathbb{R}$ holds

$$\langle \boldsymbol{x}, a\boldsymbol{y}\rangle = a\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle.$$

We have proven the case $a = 2$ in Eq. (8.7). The case $a = n \in \mathbb{N}$ is proven by induction: If $\langle \boldsymbol{x}, n\boldsymbol{y}\rangle = n\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle$, then the same relation holds for $(n+1)$, since

$$\begin{aligned}\langle \boldsymbol{x}, (n+1)\boldsymbol{y}\rangle &= \langle \boldsymbol{x}, (n\boldsymbol{y}+\boldsymbol{y})\rangle \\ &= \langle \boldsymbol{x}, n\boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle \\ &= n\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle \\ &= (n+1)\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle.\end{aligned}$$

The case $a = 1/n$ with $n \in \mathbb{N}$ comes from the relation

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \Big\langle \boldsymbol{x}, \frac{n}{n}\boldsymbol{y}\Big\rangle = n\Big\langle \boldsymbol{x}, \frac{1}{n}\boldsymbol{y}\Big\rangle \quad \Rightarrow \quad \Big\langle \boldsymbol{x}, \frac{1}{n}\boldsymbol{y}\Big\rangle = \frac{1}{n}\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle.$$

These two cases show that for any rational number $a = p/q \in \mathbb{Q}$ holds

$$\left\langle \boldsymbol{x}, \frac{p}{q}\boldsymbol{y} \right\rangle = p \left\langle \boldsymbol{x}, \frac{1}{q}\boldsymbol{y} \right\rangle = \frac{p}{q} \langle \boldsymbol{x}, \boldsymbol{y} \rangle. \tag{8.8}$$

Finally, the same property holds for all $a \in \mathbb{R}$ by the following continuity argument. Fix arbitrary vectors $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$a \mapsto f(a) = \langle \boldsymbol{x}, a\boldsymbol{y} \rangle.$$

We left as an exercise to show that f is a continuous function. Now, using Eq. (8.8) we know that this function satisfies

$$f(a) = a\, f(1) \qquad \forall\ a \in \mathbb{Q}.$$

Let $\{a_k\}_{k=1}^{\infty} \subset \mathbb{Q}$ be a sequence of rational numbers that converges to $a \in \mathbb{R}$. Since f is a continuous function we know that $\lim_{k\to\infty} f(a_k) = f(a)$. Since the sequence is constructed with rational numbers, for every element in this sequence holds

$$f(a_k) = a_k\, f(1)$$

which implies that

$$\lim_{k\to\infty} f(a_k) = \big(\lim_{k\to\infty} a_k\big)\, f(1) = a\, f(1).$$

So we have shown that for all $a \in \mathbb{R}$ holds $f(a) = a\, f(1)$, that is,

$$\langle \boldsymbol{x}, a\boldsymbol{y} \rangle = a\, \langle \boldsymbol{x}, \boldsymbol{y} \rangle.$$

Since $\boldsymbol{x}$, $\boldsymbol{y} \in V$ are fixed but arbitrary, we have shown that the function $\langle\, ,\rangle$ in linear in the second argument. We conclude that this function is an inner product on V. This establishes the Theorem in the case that V is a real vector space. □

EXAMPLE 8.1.5: Show that for $p \neq 2$ the p-norm on the vector space $\mathbb{F}^n$ introduced in Definition 8.1.2 does not satisfy the polarization identity.

SOLUTION: Consider the vectors e_1 and e_2, the first two columns of the identity matrix I_n. It is simple to compute,

$$\|\mathsf{e}_1 + \mathsf{e}_2\|_p^2 = 2^{2/p}, \quad \|\mathsf{e}_1 - \mathsf{e}_2\|_p^2 = 2^{2/p}, \quad \|\mathsf{e}_1\|_p^2 = \|\mathsf{e}_2\|_p^2 = 1$$

therefore,

$$\|\mathsf{e}_1 + \mathsf{e}_2\|_p^2 + \|\mathsf{e}_1 - \mathsf{e}_2\|_p^2 = 2^{\frac{p+2}{p}} \quad \text{and} \quad 2\big(\|\mathsf{e}_1\|_p^2 + \|\mathsf{e}_2\|_p^2\big) = 4.$$

We conclude that the p-norm satisfies the polarization identity only in the case $p = 2$. ◁

8.1.2. **Equivalent norms.** We have seen that a norm in a vector space determines a notion of distance between vectors given by the norm distance introduced in Definition 6.2.6. The notion of distance is the structure needed to define the convergence of an infinite sequence of vectors. A sequence of vectors $\{\boldsymbol{x}_i\}_{i=1}^{\infty}$ in a normed space $\big(V, \|\ \|\big)$ converges to a vector $\boldsymbol{x} \in V$ iff

$$\lim_{k\to\infty} \|\boldsymbol{x} - \boldsymbol{x}_k\| = 0,$$

that is, for every $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that for all $k > n$ holds that $\|\boldsymbol{x} - \boldsymbol{x}_k\| < \epsilon$. With the notion of convergence of a sequence it is possible to introduce concepts like the continuous and differentiable functions defined on the vector space. Therefore, the calculus can be extended from $\mathbb{R}^n$ to any normed vector space.

We have also seen that there is no unique norm in a vector space. This implies that there is no unique notion of norm distance. It is important to know whether two different norms provide the same notion of convergence of a sequence. By this we mean that every sequence that converges (diverges) with respect to one norm distance also converges (diverges) with

respect to the other norm distance. In order to answer this question it is useful the following notion.

Definition 8.1.8. *The norms* $\| \ \|_a$ *and* $\| \ \|_b$ *defined on a vector space* V *are called to be* **equivalent** *iff there exist real constants* $K \geqslant k > 0$ *such that for all* $\boldsymbol{x} \in V$ *holds*

$$k \, \|\boldsymbol{x}\|_b \leqslant \|\boldsymbol{x}\|_a \leqslant K \, \|\boldsymbol{x}\|_b.$$

It is simple to see that if two norms defined on a vector space are equivalent, then they have the same notion of convergence. What it is non-trivial to prove is that the converse also holds. Since two norms have the same notion of convergence iff they are equivalent, it is important to know whether a vector space can have non-equivalent norms. The following result addresses the case of finite dimensional vector spaces.

Theorem 8.1.9. *If* V *is a finite dimensional, then all norms defined on* V *are equivalent.*

This result says that there is a unique notion of convergence in any finite dimensional vector space. So, functions that are continuous or differentiable with respect to one norm are also continuous or differentiable with respect to any other norm. This is not the case of infinite dimensional vector spaces. It is possible to find non-equivalent norms on infinite dimensional vector spaces. Therefore, functions defined on such a vector space can be continuous with respect to one norm and discontinuous with respect to the other norm.

Proof of Theorem 8.1.9: Let $\| \ \|_a$ and $\| \ \|_b$ be two norms defined on a finite dimensional vector space V. Let $\dim V = n$, and fix a basis $\mathcal{V} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ of V. Then, any vector $\boldsymbol{x} \in V$ can be decomposed in terms of the basis vectors as

$$\boldsymbol{x} = x_1 \boldsymbol{v}_1 + \cdots + x_n \boldsymbol{v}_n.$$

Since $\| \ \|_a$ is a norm, we can obtain the following bound on $\|\boldsymbol{x}\|_a$ for every $\boldsymbol{x} \in V$ in terms of $(\sum_{i=1}^n |x_i|)$ as follows.

$$\begin{aligned} \|\boldsymbol{x}\|_a &= \|x_1 \boldsymbol{v}_1 + \cdots + x_n \boldsymbol{v}_n\|_a \\ &\leqslant |x_1| \, \|\boldsymbol{v}_1\|_a + \cdots + |x_n| \, \|\boldsymbol{v}_n\|_a \\ &\leqslant \big(|x_1| + \cdots + |x_n|\big) \, a_{\max} \quad \Rightarrow \quad \|\boldsymbol{x}\|_a \leqslant \Big(\sum_{i=1}^n |x_i|\Big) \, a_{\max}. \end{aligned}$$

where $a_{\max} = \max\{\|\boldsymbol{v}_1\|_a, \cdots, \|\boldsymbol{v}_n\|_a\}$. A lower bound can also be found as follows. Introduce the set

$$S = \Big\{ [\hat{x}_i] \in \mathbb{F}^n \ : \ \sum_{i=1}^n |\hat{x}_i| = 1 \Big\} \subset \mathbb{F}^n,$$

and then introduce the function $f_a : S \to R$ as follows

$$f_a([\hat{x}_i]) = \Big\| \sum_{i=1}^n \hat{x}_i \boldsymbol{v}_i \Big\|_a .$$

Since function f_a is a continuous and defined on a closed and bounded set of $\mathbb{F}^n$, then f_a has attains a maximum and a minimum values on S. We are here interested only in the minimum value. If $[\hat{y}_i] \in S$ is a point where f_a takes its minimum value, let let us denote

$$f_a([\hat{y}_i]) = a_{\min}.$$

Since the norm $\| \ \|_a$ is positive, we know that $a_{\min} > 0$. The existence of this minimum value of f_a implies the following bound on $\|\boldsymbol{x}\|_a$ for all $\boldsymbol{x} \in V$, namely,

$$\begin{aligned}
\|\boldsymbol{x}\|_a &= \Big\| \sum_{i=1}^{n} x_i \boldsymbol{v}_i \Big\|_a \\
&= \frac{\left(\sum_{j=1}^{n} |x_j| \right)}{\left(\sum_{j=1}^{n} |x_j| \right)} \Big\| \sum_{i=1}^{n} x_i \boldsymbol{v}_i \Big\|_a \\
&= \left(\sum_{j=1}^{n} |x_j| \right) \Big\| \sum_{i=1}^{n} \Big[\frac{x_i}{\left(\sum_{j=1}^{n} |x_j| \right)} \Big] \boldsymbol{v}_i \Big\|_a \\
&= \left(\sum_{j=1}^{n} |x_j| \right) \Big\| \sum_{i=1}^{n} \hat{x}_i \, \boldsymbol{v}_i \Big\|_a \\
&= \left(\sum_{j=1}^{n} |x_j| \right) f_a([\hat{x}_i]) \quad \Rightarrow \quad \|\boldsymbol{x}\|_a \geqslant \left(\sum_{j=1}^{n} |x_j| \right) a_{\min}.
\end{aligned}$$

Summarizing, we have found real numbers $a_{\max} \geqslant a_{\min} > 0$ such that the following inequality holds for all $\boldsymbol{x} \in V$,

$$a_{\min} \left(\sum_{j=1}^{n} |x_j| \right) \leqslant \|\boldsymbol{x}\|_a \leqslant \left(\sum_{j=1}^{n} |x_j| \right) a_{\max}.$$

Since no special property of norm $\| \ \|_a$ has been used, the same type of inequality holds for norm $\| \ \|_b$, that is, there exist real numbers $b_{\max} \geqslant b_{\min} > 0$ such that the following inequality holds for all $\boldsymbol{x} \in B$,

$$b_{\min} \left(\sum_{j=1}^{n} |x_j| \right) \leqslant \|\boldsymbol{x}\|_b \leqslant \left(\sum_{j=1}^{n} |x_j| \right) b_{\max}.$$

These inequalities imply that norms $\| \ \|_a$ and $\| \ \|_b$ are equivalent, since

$$\|\boldsymbol{x}\|_b \, \frac{a_{\min}}{b_{\max}} \leqslant \frac{b_{\max}}{b_{\max}} a_{\min} \left(\sum_{j=1}^{n} |x_j| \right) \leqslant \|\boldsymbol{x}\|_a \leqslant \left(\sum_{j=1}^{n} |x_j| \right) a_{\max} \frac{b_{\min}}{b_{\min}} \leqslant \frac{a_{\max}}{b_{\min}} \, \|\boldsymbol{x}\|_b.$$

(Start reading the inequality form the center, at $\|\boldsymbol{x}\|_a$, and first see the inequalities to the right; then go to the center again and read the inequalities to the left.) Denoting $k = a_{\min}/b_{\max}$ and $K = a_{\max}/b_{\min}$, we have obtained that

$$k \, \|\boldsymbol{x}\|_b \leqslant \|\boldsymbol{x}\|_a \leqslant K \, \|\boldsymbol{x}\|_b \qquad \forall \ \boldsymbol{x} \in V.$$

This establishes the Theorem. □

8.1.3. **Exercises.**

8.1.1.- Consider the vector space $\mathbb{C}^4$ and for $p = 1, 2, \infty$ find the p-norms of the vectors

$$\mathsf{x} = \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix}, \quad \mathsf{y} = \begin{bmatrix} 1+i \\ 1-i \\ 1 \\ 4i \end{bmatrix}.$$

8.1.2.- Determine which of the following functions $\| \ \| : \mathbb{R}^2 \to \mathbb{R}$ defines a norm on $\mathbb{R}^2$. We denote $\mathsf{x} = [x_i]$ the components of x in a standard basis of $\mathbb{R}^2$. Justify your answers.

(a) $\|\mathsf{x}\| = |x_1|$;
(b) $\|\mathsf{x}\| = |x_1 + x_2|$;
(c) $\|\mathsf{x}\| = |x_1|^2 + |x_2|^2$;
(d) $\|\mathsf{x}\| = 2|x_1| + 3|x_2|$.

8.1.3.- True or false? Justify your answer: If $\| \ \|_a$ and $\| \ \|_b$ are two norms on a vector space V, then $\| \ \|$ defined as

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_a + \|\boldsymbol{x}\|_b \quad \forall\ \boldsymbol{x} \in V$$

is also a norm on V.

8.1.4.- Consider the space $\mathbb{P}_2([0,1])$ with the p-norm

$$\|\boldsymbol{q}\|_p = \Big(\int_0^1 |\boldsymbol{q}(x)|^p \, dx \Big)^{\frac{1}{p}},$$

for $p \in [1, \infty)$. Find the p-norm of the vector $\boldsymbol{q}(x) = -3x^2$. Also find the supremum norm of $\boldsymbol{q}$, defined as

$$\|\boldsymbol{q}\|_\infty = \max_{x \in [0,1]} |\boldsymbol{q}(x)|.$$

8.2. Operator norms

We have seen that the space of all linear transformations $L(V,W)$ between the vector spaces V and W is itself a vector space. As in any vector space, it is possible to define norm functions $\|\ \| : L(V,W) \to \mathbb{R}$ on the vector space $L(V,W)$. For example, We will see that in the case where $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, one of such norms is the Frobenius norm, which is the inner product norm corresponding to the Frobenius inner product introduced in Section 6.2. More precisely, the **Frobenius norm** is given by

$$\|\mathsf{A}\|_F = \sqrt{\langle \mathsf{A}, \mathsf{A}\rangle_F} \qquad \forall\ \mathsf{A} \in \mathbb{F}^{m,n}.$$

It is simple to see that

$$\|\mathsf{A}\|_F = \sqrt{\mathrm{tr}\left(\mathsf{A}^*\mathsf{A}\right)} = \Big(\sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|^2\Big)^{\frac{1}{2}}.$$

However, in the case where V and W are normed spaces with norms $\|\ \|_{\mathrm{v}}$ and $\|\ \|_{\mathrm{w}}$, there exists a particular norm on $L(V,W)$ which is induced from the norms on V and W. This induced norm can be defined on elements $L(V,W)$ by recalling that the element $\boldsymbol{T} \in L(V,W)$ as a function $\boldsymbol{T} : V \to W$.

Definition 8.2.1. *Let $\big(V, \|\ \|_{\mathrm{v}}\big)$ and $\big(W, \|\ \|_{\mathrm{w}}\big)$ be finite dimensional normed spaces. The* **induced norm** *on the space $L(V,W)$ of all linear transformations $\boldsymbol{T} : V \to W$ is a function $\|\ \| : L(V,W) \to \mathbb{R}$ given by*

$$\|\boldsymbol{T}\| = \max_{\|\boldsymbol{x}\|_{\mathrm{v}}=1} \|\boldsymbol{T}(\boldsymbol{x})\|_{\mathrm{w}}.$$

In the particular case that $V = W$ and $\|\ \|_{\mathrm{v}} = \|\ \|_{\mathrm{w}} = \|\ \|$, the induced norm on $L(V)$ is called the **operator norm**, *and is given by*

$$\|\boldsymbol{T}\| = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{T}(\boldsymbol{x})\|.$$

The definition above says that given arbitrary norms on the vector spaces V and W, they induce a norm on the vector space $L(V,W)$ of linear transformations $\boldsymbol{T} : V \to W$. A particular case of this definition is when $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, we fix standard bases on V and W, and we introduce p-norms on these spaces. So, the normed spaces $\big(\mathbb{F}^n, \|\ \|_p\big)$ and $\big(\mathbb{F}^m, \|\ \|_q\big)$ induce a norm on the vector space $\mathbb{F}^{m,n}$ as follows.

Definition 8.2.2. *Consider the normed spaces $\big(\mathbb{F}^n, \|\ \|_p\big)$ and $\big(\mathbb{F}^m, \|\ \|_q\big)$, with $p, q \in [1,\infty]$. The* **induced (p,q)-norm** *on the space $L(\mathbb{F}^n, \mathbb{F}^m)$ is the function $\|\ \|_{p,q} : L(\mathbb{F}^n, \mathbb{F}^m) \to \mathbb{R}$ given by*

$$\|\boldsymbol{T}\|_{p,q} = \max_{\|\mathsf{x}\|_p=1} \|\boldsymbol{T}(\mathsf{x})\|_q \qquad \forall\ \boldsymbol{T} \in L(\mathbb{F}^n, \mathbb{F}^m).$$

In the particular case of $p = q$ we denote $\|\ \|_{p,p} = \|\ \|_p$. In the case $p = q$ and $n = m$ the induced norm on $L(\mathbb{F}^n)$ is called the **p-operator norm** *and is given by*

$$\|\boldsymbol{T}\|_p = \max_{\|\mathsf{x}\|_p=1} \|\boldsymbol{T}(\mathsf{x})\|_p \qquad \forall\ \boldsymbol{T} \in L(\mathbb{F}^n).$$

In the case $p = q$ above we use the same notation $\|\ \|_p$ for the p-norm on $\mathbb{F}^n$, the (p,p)-norm on $L(\mathbb{F}^n, \mathbb{F}^m)$ and the p-operator nor in $L(\mathbb{F}^n)$. The context should help to decide which norm we use on every particular situation.

Example 8.2.1: Consider the particular case $V = W = \mathbb{R}^2$ with standard ordered bases and the $p = 2$ norms in both, V and W. In this case, the space $L(\mathbb{R}^2)$ can be identified with

the space $\mathbb{R}^{2,2}$ of all 2×2 matrices. The induced norm on $\mathbb{R}^{2,2}$, denoted as $\| \ \|_2$ and called the operator norm, is the following:

$$\|\mathsf{A}\|_2 = \max_{\|\mathsf{x}\|_2=1} \|\mathsf{A}\mathsf{x}\|_2, \qquad \forall\ \mathsf{A} \in \mathbb{R}^{2,2}.$$

The meaning of this norm is deeply related with the interpretation of the matrix A as a linear operator $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$. Suppose that the action of the operator A on the unit circle (B_2 in the notation of Example 8.1.2) is given in Fig. 55. Then the value of the operator norm $\|\mathsf{A}\|_2$ is the size measured in the 2-norm of the maximum deformation by A of the unit circle.

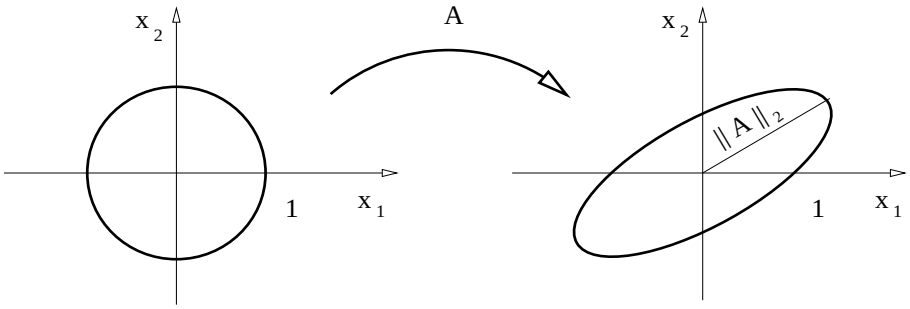

FIGURE 55. Geometrical meaning of the operator norm on $\mathbb{R}^{2,2}$ induced by the 2-norm on $\mathbb{R}^2$.

◁

EXAMPLE 8.2.2: Consider the normed space $\big(\mathbb{R}^2, \| \ \|_2\big)$, and let $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$ be the matrix

$$\mathsf{A} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \quad \text{with} \quad |A_{11}| \neq |A_{22}|.$$

Find the 2-operator norm induced on A.

SOLUTION: Since the norm on $\mathbb{R}^2$ is the 2-norm, the induced norm on A is given by

$$\|\mathsf{A}\|_2 = \max_{\|\mathsf{x}\|_2=1} \|\mathsf{A}\mathsf{x}\|_2.$$

We need to find the maximum of $\|\mathsf{A}\mathsf{x}\|_2$ among all x subject to the constraint $\|\mathsf{x}\|_2 = 1$. So, this is a constrained maximization problem, that is, a maxima-minima problem where the variable x is restricted by a constraint equation. In general, this type of problems can be solved using the Lagrange multipliers method. Since this example is simple enough, we solve it in a simpler way. We first solve the constraint equation on x, and we then find the maxima of $\|\mathsf{A}\mathsf{x}\|_2$ among these solutions only. The general solution of the equation

$$\|\mathsf{x}\|_2 = 1 \quad \Leftrightarrow \quad (x_1)^2 + x_2)^2 = 1$$

is given by

$$\mathsf{x}(\theta) = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} \quad \text{with} \quad \theta \in [0, 2\pi).$$

Introduce this general solution into $\|\mathsf{A}\mathsf{x}\|_2$ we obtain

$$\|\mathsf{A}\mathsf{x}\|_2 = \sqrt{(A_{11})^2 \cos^2(\theta) + (A_{22})^2 \sin^2(\theta)}.$$

Since the maximum in θ of the function $\|\mathsf{A}\mathsf{x}\|_2$ is the same as the maximum of $\|\mathsf{A}\mathsf{x}\|_2^2$, we need to find the maximum on θ of the function $f(\theta) = \|\mathsf{A}\mathsf{x}(\theta)\|_2^2$, that is,

$$f(\theta) = (A_{11})^2 \cos^2(\theta) + (A_{22})^2 \sin^2(\theta).$$

The solution is simple, find θ solutions of

$$f'(\theta) = \frac{df}{d\theta}(\theta) = 0 \quad \Rightarrow \quad 2\big[-(A_{11})^2 + (A_{22})^2\big] \sin(\theta)\cos(\theta) = 0.$$

Since we assumed that $|A_{11}| \neq |A_{22}|$, then

$$\sin(\theta) = 0 \quad \Rightarrow \quad \theta_1 = 0, \quad \theta_2 = \pi,$$

$$\cos(\theta) = 0 \quad \Rightarrow \quad \theta_3 = \frac{\pi}{2}, \quad \theta_4 = \frac{3\pi}{2}.$$

We obtained four solutions for θ. We evaluate f at these solutions,

$$f(0) = f(\pi) = (A_{11})^2, \quad f\big(\frac{\pi}{2}\big) = f\big(\frac{3\pi}{2}\big) = (A_{22})^2.$$

Recalling that $\|\mathsf{A}\mathsf{x}(\theta)\|_2 = \sqrt{f(\theta)}$, we obtain

$$\|\mathsf{A}\|_2 = \max\big\{|A_{11}|, |A_{22}|\big\}.$$

◁

It was mentioned in Example 8.2.2 above that finding the operator norm requires solving a constrained maximization problem. In the case that the operator norm is induced from a dot product norm, the constrained maximization problem can be solved in an explicit form.

Proposition 8.2.3. *Consider the vector spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ with inner product given by the dot products, inner product norms $\|\ \|_2$, and the vector space $L(\mathbb{R}^n, \mathbb{R}^m)$ with the induced 2-norm.*

$$\|\mathsf{A}\|_2 = \max_{\|\mathsf{x}\|_2 = 1} \|\mathsf{A}\mathsf{x}\|_2 \quad \forall\ \mathsf{A} \in L(\mathbb{R}^n, \mathbb{R}^m).$$

Introduce the scalars $\lambda_i \in \mathbb{R}$, with $i = 1, \cdots, k \leqslant n$, as all the roots of the polynomial

$$p(\lambda) = \det\big(\mathsf{A}^T\mathsf{A} - \lambda\, \mathsf{I}_n\big).$$

Then, all scalars λ_i are non-negative real numbers and the induced 2-norm of the transformation A is given by

$$\|\mathsf{A}\|_2 = \max\big\{\lambda_1, \cdots, \lambda_k\big\}.$$

Proof of Proposition 8.2.3: Introduce the functions $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ as follows,

$$f(\mathsf{x}) = \|\mathsf{A}\mathsf{x}\|_2^2 = \big(\mathsf{A}\mathsf{x}\big) \cdot \big(\mathsf{A}\mathsf{x}\big) = \mathsf{x}^T\mathsf{A}^T\mathsf{A}\mathsf{x}, \quad g(\mathsf{x}) = \|\mathsf{x}\|_2^2 = \mathsf{x} \cdot \mathsf{x} = \mathsf{x}^T\mathsf{x}.$$

To find the induced norm of $\boldsymbol{T}$ is then equivalent to solve the constrained maximization problem: Find the maximum of $f(\mathsf{x})$ for $\mathsf{x} \in \mathbb{R}^n$ subject to the constraint $g(\mathsf{x}) = 1$. The vectors x that provide solutions to the constrained maximization problem must we solutions of the Euler-Lagrange equations

$$\nabla f = \lambda\, \nabla g \tag{8.9}$$

where $\lambda \in \mathbb{R}$, and we introduced the gradient row vectors

$$\nabla f = \left[\frac{\partial f}{\partial x_1} \quad \cdots \quad \frac{\partial f}{\partial x_n}\right], \quad \nabla f = \left[\frac{\partial g}{\partial x_1} \quad \cdots \quad \frac{\partial g}{\partial x_n}\right].$$

In order to understand why the solution x must satisfy the Euler-Lagrange equations above we need to recall two properties of the gradient vector. First, the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a vector that determines the direction on $\mathbb{R}^n$ where f has the maximum increase. Second, which is deeply related to the first property, the gradient vector of a

function is perpendicular to the surfaces where the function has a constant value. The surfaces of constant value of a function are called level surfaces of the function. Therefore, the function f has a maximum or minimum value at x on the constraint level surface $g = 1$ if ∇f is perpendicular to the level surface $g = 1$ at that x. (Proof: Suppose that at a particular $\boldsymbol{x}$ on the constraint surface $g = 1$ the projection of ∇f onto the constraint surface is nonzero; then the values of f increase along that direction on the constraint surface; this means that f does not attain a maximum value at that x on $g = 1$.) We conclude that both gradients ∇f and ∇g are parallel, which is precisely what Eq. (8.9) says. In our particular problem we obtain for f and g the following:

$$f(\mathsf{x}) = \mathsf{x}^T\mathsf{A}^T\mathsf{A}\mathsf{x} \quad \Rightarrow \quad \nabla f(\mathsf{x}) = 2\mathsf{x}^T\mathsf{A}^T\mathsf{A},$$

$$g(\mathsf{x}) = \mathsf{x}^T\mathsf{x} \quad \Rightarrow \quad \nabla g(\mathsf{x}) = 2\mathsf{x}^T.$$

We must look for $\mathsf{x} \neq \mathsf{0}$ solution of the equation

$$\mathsf{x}^T\mathsf{A}^T\mathsf{A} = \lambda\,\mathsf{x}^T \quad \Leftrightarrow \quad \mathsf{A}^T\mathsf{A}\mathsf{x} = \lambda\,\mathsf{x},$$

where the condition $\mathsf{x} \neq \mathsf{0}$ comes from $\|\mathsf{x}\|_2 = 1$. Therefore, λ must not be any scalar but the precise scalar or scalars such that the matrix $(\mathsf{A}^T\mathsf{A} - \lambda\,\mathsf{I}_n)$ is not invertible. An equivalent condition is that

$$p(\lambda) = \det\big(\mathsf{A}^T\mathsf{A} - \lambda\,\mathsf{I}_n\big) = 0.$$

The function p is a polynomial of degree n in λ so it has at most n real roots. Let us denote these roots by $\lambda_1, \cdots, \lambda_k$, with $1 \leqslant k \leqslant n$. For each of these values λ_i the matrix $\mathsf{A}^T\mathsf{A} - \lambda_i\,\mathsf{I}_n$ is not invertible, so $N(\mathsf{A}^T\mathsf{A} - \lambda_i\,\mathsf{I}_n)$ is non-trivial. Let x_i be any element in $N(\mathsf{A}^T\mathsf{A} - \lambda_i\,\mathsf{I}_n)$, that is,

$$\mathsf{A}^T\mathsf{A}\mathsf{x}_i = \lambda_i\,\mathsf{x}_i, \qquad i = 1, \cdots, k.$$

At this point it is not difficult to see that $\lambda_i \geqslant 0$ for $i = 0, \cdots, k$. Indeed, multiply the equation above by x_i^T, that is,

$$\mathsf{x}_i^T\mathsf{A}^T\mathsf{A}\mathsf{x}_i = \lambda_i\,\mathsf{x}_i^T\mathsf{x}_i.$$

Since both $\mathsf{x}_i^T\mathsf{A}^T\mathsf{A}\mathsf{x}_i$ and $\mathsf{x}_i^T\mathsf{x}_i$ are non-negative numbers, so is λ_i. Returning to x_i, only these vectors are the candidates to find a solution to our maximization problem, and f has the values

$$f(\mathsf{x}_i) = \mathsf{x}_i^T\mathsf{A}^T\mathsf{A}\mathsf{x}_i = \lambda_i\,\mathsf{x}_i^T\mathsf{x}_i = \lambda_i \quad \Rightarrow \quad f(\mathsf{x}_i) = \lambda_i, \qquad i = 1, \cdots, k.$$

The induced 2-norm of A is the maximum of these scalar λ_i. This establishes the Proposition. □

EXAMPLE 8.2.3: Consider the vector spaces $\mathbb{R}^3$ and $\mathbb{R}^2$ with the dot product. Find the induced 2-norm of the 2×3 matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

SOLUTION: Following the Proposition 8.2.3 the value of the induced norm $\|A\|_2$ is the maximum of the λ_i roots of the polynomial

$$p(\lambda) = \det\big(\mathsf{A}^T\mathsf{A} - \lambda\,\mathsf{I}_3\big) = 0.$$

We start computing

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 2 \end{bmatrix}.$$

The next step is to find the polynomial

$$p(\lambda) = \begin{vmatrix} (1-\lambda) & 0 & 1 \\ 0 & (1-\lambda) & -1 \\ 1 & -1 & (2-\lambda) \end{vmatrix} = (1-\lambda)\big[(1-\lambda)(2-\lambda) - 1\big] + (1-\lambda),$$

therefore, we obtain $p(\lambda) = -\lambda(\lambda-1)(\lambda-3)$. We have three roots

$$\lambda_1 = 0, \quad \lambda_2 = 1, \quad \lambda_3 = 3.$$

We then conclude that the induced norm of A is given by

$$\|\mathsf{A}\|_2 = 3.$$

◁

In the case that the norm in a normed space is not an inner product norm the induced norm on linear operators is not simple to evaluate. The constrained maximization problem is in general complicated to solve. Two particular cases can be solved explicitly though, when the operator norm is induced from the p-norms with $p = 1$ and $p = \infty$.

Proposition 8.2.4. *Consider the normed spaces $\big(\mathbb{R}^n, \|\ \|_p\big)$ and $\big(\mathbb{R}^m, \|\ \|_p\big)$ and the vector space $L(\mathbb{R}^n, \mathbb{R}^m)$ with the induced p-norm.*

$$\|\mathsf{A}\|_p = \max_{\|\mathsf{x}\|_p = 1} \|\mathsf{A}(\mathsf{x})\|_p \quad \forall\ \mathsf{A} \in L(\mathbb{R}^n, \mathbb{R}^m).$$

If $p = 1$ or $p = \infty$, then the following formulas hold, respectively,

$$\|\mathsf{A}\|_1 = \max_{j\in\{1,\cdots,n\}} \sum_{i=1}^{m} |A_{ij}|, \qquad \|\mathsf{A}\|_\infty = \max_{i\in\{1,\cdots,m\}} \sum_{j=1}^{n} |A_{ij}|.$$

Proof of proposition 8.2.4: From the definition of the induced p-norm for $p = 1$,

$$\|\mathsf{A}\|_1 = \max_{\|\mathsf{x}\|_1 = 1} \|\mathsf{A}\mathsf{x}\|_1.$$

From the p-norm on $\mathbb{R}^m$ we know that $\|\mathsf{A}\mathsf{x}\|_1 = \sum_{i=1}^{m} \Big| \sum_{j=1}^{n} A_{ij} x_j \Big|$, therefore,

$$\|\mathsf{A}\mathsf{x}\|_1 \leqslant \sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|\,|x_j| = \sum_{j=1}^{n} |x_j| \sum_{i=1}^{m} |A_{ij}| \leqslant \Big(\sum_{j=1}^{n} |x_j|\Big) \max_{j\in\{1,\cdots,n\}} \Big(\sum_{i=1}^{m} |A_{ij}|\Big);$$

and introducing the condition $\|\mathsf{x}\|_1 = 1$ we obtain the inequality

$$\|\mathsf{A}\mathsf{x}\|_1 \leqslant \max_{j\in\{1,\cdots,n\}} \sum_{i=1}^{m} |A_{ij}|.$$

Recalling the column vector notation $\mathsf{A} = \big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big]$, we notice that $\|\mathsf{A}_{:j}\|_1 = \sum_{i=1}^{m} |A_{ij}|$, so the inequality above can be expressed as

$$\|\mathsf{A}\mathsf{x}\|_1 \leqslant \max_{j\in\{1,\cdots,n\}} \|\mathsf{A}_{:j}\|_1.$$

Since the left hand side is independent of x, the inequality also holds for the maximum in $\|\mathsf{x}\|_1 = 1$, that is,

$$\|\mathsf{A}\|_1 \leqslant \max_{j\in\{1,\cdots,n\}} \|\mathsf{A}_{:j}\|_1.$$

It is now clear that the equality has to be achieved, since the $\|\mathsf{A}\mathsf{x}\|_1$ in the case $\mathsf{x} = \mathsf{e}_j$, with e_j a standard basis vector, takes the value $\|\mathsf{A}\mathsf{e}_j\|_1 = \|\mathsf{A}_{:j}\|_1$. Therefore,

$$\|\mathsf{A}\|_1 = \max_{j\in\{1,\cdots,n\}} \|\mathsf{A}_{:j}\|_1.$$

The second part of Proposition 8.2.4 can be proven as follows. From the definition of the induced p-norm for $p = \infty$ we know that

$$\|\mathsf{A}\|_\infty = \max_{\|\mathsf{x}\|_\infty = 1} \|\mathsf{A}\mathsf{x}\|_\infty.$$

From the p-norm on $\mathbb{R}^m$ we see

$$\|\mathsf{A}\mathsf{x}\|_\infty = \max_{i\in\{1,\cdots,m\}} \Big|\sum_{j=1}^{n} A_{ij}x_j\Big| \leqslant \max_{i\in\{1,\cdots,m\}} \sum_{j=1}^{n} |A_{ij}|\,|x_j| \leqslant \max_{i\in\{1,\cdots,m\}} \sum_{j=1}^{n} |A_{ij}|.$$

Since the far right hand side does not depend on x, the inequality must hold for the maximum in x with $\|\mathsf{x}\|_\infty = 1$, so we conclude that

$$\|A\|_\infty \leqslant \max_{i\in\{1,\cdots,m\}} \sum_{j=1}^{n} |A_{ij}|.$$

As in the previous $p = 1$ case, the value on the right hand side above is achieved for $\|\mathsf{A}\mathsf{x}\|_\infty$ in the case of x with components ± 1 depending on the sign of A_{ij}. More precisely, choose x as follows,

$$x_j = \begin{cases} 1 & \text{if} \quad A_{ij} \geqslant 0, \\ -1 & \text{if} \quad A_{ij} < 0, \end{cases} \quad \Rightarrow \quad \sum_{j=1}^{n} A_{ij}x_j = \sum_{j=1}^{n} |A_{ij}|, \quad i = 1, \cdots, m.$$

Therefore, for that x we have $\|\mathsf{x}\|_\infty = 1$ and $\|\mathsf{A}\mathsf{x}\|_\infty = \max_{i\in\{1,\cdots,m\}} \sum_{j=1}^{n} |A_{ij}|$. So, we conclude

$$\|\mathsf{A}\|_\infty = \max_{i\in\{1,\cdots,m\}} \sum_{j=1}^{n} |A_{ij}|.$$

This establishes the Proposition □

EXAMPLE 8.2.4: Find the induced p-norm, where $p = 1, \infty$, for the 2×3 matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 1 & -1 \end{bmatrix}.$$

SOLUTION: Proposition 8.2.4 says that $\|\mathsf{A}\|_1$ is the largest absolute value sum of components among columns of A, while $\|\mathsf{A}\|_\infty$ is the largest absolute value sum among rows of A. In the first case we have:

$$\|\mathsf{A}\|_1 = \max_{j=1,2,3} \sum_{i=1}^{2} |A_{ij}|;$$

since

$$\sum_{i=1}^{2} |A_{i1}| = 3, \qquad \sum_{i=1}^{2} |A_{i2}| = 5, \qquad \sum_{i=1}^{2} |A_{i3}| = 2,$$

therefore, $\|\mathsf{A}\|_1 = 5$. In the second case e have:

$$\|\mathsf{A}\|_\infty = \max_{i=1,2} \sum_{j=1}^{2} |A_{ij}|;$$

since

$$\sum_{j=1}^{3} |A_{1j}| = 6, \qquad \sum_{j=1}^{3} |A_{2j}| = 4,$$

therefore, $\|\mathsf{A}\|_\infty = 6$. ◁

8.2.1. **Exercises.**

8.2.1.- Evaluate the induced p-norm, where $p = 1, 2, \infty$, for the matrices

$$\mathsf{A} = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

8.2.2.- In the normed space $\big(\mathbb{R}^2, \| \ \|_2\big)$, find the induced norm of $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$

$$\mathsf{A} = \frac{1}{\sqrt{3}} \begin{bmatrix} 3 & -1 \\ 0 & \sqrt{8} \end{bmatrix}.$$

8.2.3.- Consider the space $\big(\mathbb{F}^n, \| \ \|_p\big)$ and the space $\mathbb{F}^{n,n}$ with the induced norm $\| \ \|_p$. Prove that for all $\mathsf{A}, \mathsf{B} \in \mathbb{F}^{n,n}$ and all $\mathsf{x} \in \mathbb{F}^n$ holds

(a) $\|\mathsf{Ax}\|_p \leqslant \|\mathsf{A}\|_p \, \|\mathsf{x}\|_p$;

(b) $\|\mathsf{AB}\|_p \leqslant \|\mathsf{A}\|_p \, \|\mathsf{B}\|_p$.

8.3. Condition numbers

In Sect. 1.5 we discussed several types of approximations errors that appear when solving an $m \times n$ linear system in a floating-point number set using rounding. In this Section we discuss a particular type of square linear systems with unique solutions that are greatly affected by small changes in the coefficients of their augmented matrices. We will call such systems ill-conditioned. When an ill-conditioned $n \times n$ linear system is solved on a floating-point number set using rounding, a small rounding error in the coefficients of the system may produce an approximate solution that differs significantly from the exact solution.

Definition 8.3.1. *An $n \times n$ linear system having a unique solution is* **ill-conditioned** *iff a 1% perturbation in a coefficient of its augmented matrix produces a perturbed linear system still having a unique solution which differs from the unperturbed solution in a 100% or more.*

We remark that the choice of the values 1% and 100% in the above definition is not a standard choice in the literature. While these values may change on different books, the idea behind the definition of an ill-conditioned system is still the same, that a small change in a coefficient of the linear system produces a big change in its solution.

We also remark that our definition of an ill-conditioned system applies only to square linear system having a unique solution, and such that the perturbed linear system also has a unique solution. The concept of ill-conditioned system can be generalized to other linear systems, but we do not study those cases here.

The following example gives some insight to understand what causes a 2×2 system to be ill-conditioned.

Example 8.3.1: It is not difficult to understand when a 2×2 linear system is ill-conditioned. The solution of a 2×2 linear system can be thought as the intersection of two lines on the plane, where each line represents the solution of each equation of the system. A 2×2 linear system is ill-conditioned when these intersecting lines are almost parallel. Then, a small change in the coefficients of the system produces a small change in the lines representing the solution of each equation. Since the lines are almost parallel, this small change on the lines may produce a large change of the intersection point. This situation is sketched on Fig. 56.

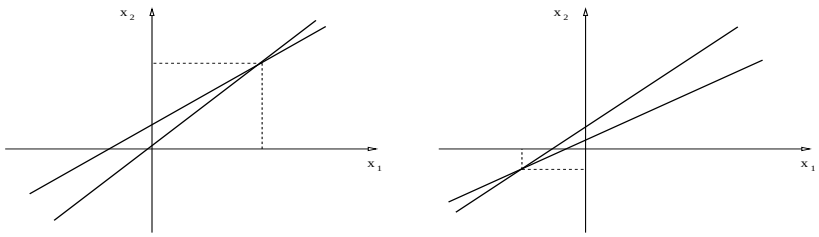

Figure 56. The intersection point of almost parallel lines represents a solution of an ill-conditioned 2×2 linear system. A small perturbation in the coefficients of the system produces a small change in the lines, which in turn produces a large change in the solution of the system.

◁

Example 8.3.2: Show that the following 2×2 linear system is ill-conditioned:

$$0.835\,x_1 + 0.667\,x_2 = 0.168,$$
$$0.333\,x_1 + 0.266\,x_2 = 0.067.$$

SOLUTION: We first note that the solution to the linear system above is given by

$$x_1 = 1, \qquad x_2 = -1.$$

In order to show that the system above is ill-conditioned we only need to find a coefficient in the system such that a small change in that coefficient produces a large change in the solution. Consider the following change on the second source coefficient:

$$0.067 \to 0.066.$$

One can check that the solution to the new system

$$0.835\, x_1 + 0.667\, x_2 = 0.168,$$
$$0.333\, x_1 + 0.266\, x_2 = 0.066,$$

is given by

$$x_1 = -666, \qquad x_2 = 834.$$

We then conclude that the system is ill-conditioned. ◁

Summarizing, solving a linear system in a floating-point number set using rounding introduces approximation errors in the coefficients of the system. The modified Gauss-Jordan method on the floating point number set also introduces approximation errors in the solution of the system. Both type of approximation errors can be controlled, that is, be kept small, choosing a particular scheme of Gauss operations; for example partial pivoting and complete pivoting. However, if the original linear system we solve is ill-conditioned, then even very small approximation errors in the system coefficients and in the Gauss operations may result in a huge error in the solutions. Therefore, it is important to prevent solving ill-conditioned systems when approximation errors are unavoidable. How handle such situations is one important research area in numerical analysis.

8.3.1. **Exercises.**

8.3.1.- Consider the ill-conditioned system from Example 8.3.2,

$$0.835\,x_1 + 0.667\,x_2 = 0.168,$$
$$0.333\,x_1 + 0.266\,x_2 = 0.067.$$

(a) Solve this system in $\mathbb{F}_{5,10,6}$ without scaling, partial or complete pivoting.

(b) Solve this system in $\mathbb{F}_{6,10,6}$ without scaling, partial or complete pivoting.

(c) Compare the results found in parts (a) and (b) with the result in $\mathbb{R}$.

8.3.2.- Perturb the ill-conditioned system given in Exercise **8.3.1** as follows,

$$0.835\,x_1 + 0.667\,x_2 = 0.1669995,$$
$$0.333\,x_1 + 0.266\,x_2 = 0.066601.$$

Find the solution of this system in $\mathbb{R}$ and compare it with the solution in Exercise **8.3.1**.

8.3.3.- Find the solution in $\mathbb{R}$ of the following system

$$8\,x_1 + 5\,x_2 + 2\,x_3 = 15,$$
$$21\,x_1 + 19\,x_2 + 16\,x_3 = 56,$$
$$39\,x_1 + 48\,x_2 + 53\,x_3 = 140.$$

Then, change 15 to 14 in the first equation and solve it again in $\mathbb{R}$. Is this system ill-conditioned?

Chapter 9. Spectral decomposition

In Sect.5.4 we discussed the matrix representation of linear transformations defined on finite dimensional vector spaces. We saw that this representation is basis-dependent. The matrix of a linear transformation can be complicated in one basis and simple in another basis. In the case of linear operators sometimes there exists a special basis where the operator matrix is the simplest possible: A diagonal matrix. In this Chapter we study those linear operators on finite dimensional vector spaces having a diagonal matrix representation, called normal operators. We start introducing the eigenvalues and eigenvectors of a linear operator. Later on we present the main result, the Spectral Theorem for normal operators. We use this result to define functions of operators. We finally mention how to apply these ideas to find solutions to a linear system of ordinary differential equations.

9.1. Eigenvalues and eigenvectors

9.1.1. **Main definitions.** Sometimes a linear operator has the following property: The image under the operator of a particular line in the vector space is again the same line. In that case we call the line special, eigen in German. Any non-zero vector in that line is also special and is called an eigenvector.

Definition 9.1.1. *Let V be a finite dimensional vector space over the field $\mathbb{F}$. The scalar $\lambda \in \mathbb{F}$ and the non-zero vector $\boldsymbol{x} \in V$ are called **eigenvalue** and **eigenvector** of the linear operator $\boldsymbol{T} \in L(V)$ iff holds*

$$\boldsymbol{T}(\boldsymbol{x}) = \lambda\, \boldsymbol{x}. \tag{9.1}$$

*The set $\sigma_T \subset \mathbb{F}$ of all eigenvalues of the operator $\boldsymbol{T}$ is called the **spectrum** of $\boldsymbol{T}$. The subspace $E_\lambda = N(\boldsymbol{T} - \lambda\, \boldsymbol{I}) \subset V$, the null space of the operator $(\boldsymbol{T} - \lambda\, \boldsymbol{I})$, is called the **eigenspace** of $\boldsymbol{T}$ corresponding to λ.*

An eigenvector of a linear operator $\boldsymbol{T} : V \to V$ is a vector that remain invariant except by scaling under the action of $\boldsymbol{T}$. The change in scaling of the eigenvector $\boldsymbol{x}$ under $\boldsymbol{T}$ determines the eigenvalue λ. Since the operator $\boldsymbol{T}$ is linear, given an eigenvector $\boldsymbol{x}$ and any non-zero scalar $a \in \mathbb{F}$, the vector $a\boldsymbol{x}$ is also an eigenvector. (Proof: $\boldsymbol{T}(a\boldsymbol{x}) = a\,\boldsymbol{T}(\boldsymbol{x}) = \lambda(a\boldsymbol{x})$.) The elements of the eigenspace E_λ are all eigenvectors with eigenvalue λ and the zero vector. Indeed, a vector $\boldsymbol{x} \in E_\lambda$ iff holds $(\boldsymbol{T} - \lambda\, \boldsymbol{I})(\boldsymbol{x}) = \boldsymbol{0}$, where $\boldsymbol{I} \in L(V)$ is the identity operator, and this equation implies that $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{T}(\boldsymbol{x}) = \lambda\, \boldsymbol{x}$.

Eigenvalues and eigenvectors are notions defined on an operator, independently of any basis on the vector space. However, given a basis in a vector space, the eigenvalue-eigenvector equation can be expressed in terms of a matrix-vector product. This is summarized in the following result.

Theorem 9.1.2. *If $\mathsf{T}_{vv} \in F^{n,n}$ is the matrix of a linear operator $\boldsymbol{T} \in L(V)$ in any ordered basis $\mathcal{V} \subset V$, then the eigenvalue $\lambda \in \mathbb{F}$ and eigenvector components $\mathsf{x}_v \in \mathbb{F}^n$ of the operator $\boldsymbol{T}$ satisfy the eigenvalue-eigenvector equation*

$$\mathsf{T}_{vv}\mathsf{x}_v = \lambda\, \mathsf{x}_v. \tag{9.2}$$

The eigenvalues and eigenvectors of a linear operator are the eigenvalues and eigenvectors of any matrix representation of the operator on any ordered basis of the vector space.

Proof of Theorem 9.1.2: The eigenvalue-eigenvector equation in (9.1) written in an ordered basis $\mathcal{V} \subset V$ is given by

$$\big[\boldsymbol{T}(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n)\big]_v = \lambda[\boldsymbol{x}]_v \quad \Leftrightarrow \quad \big[[\boldsymbol{T}(\boldsymbol{v}_1)]_v, \cdots, [\boldsymbol{T}(\boldsymbol{v}_n)]_v\big]\,[\boldsymbol{x}]_v = \lambda[\boldsymbol{x}]_v,$$

that is, we obtain Eq. (9.2). It is simple to see that Eq. (9.2) is invariant under similarity transformations of the matrix T_{vv}. This property says that the components equation in (9.2) looks the same in any basis. Indeed, given any other ordered basis $\tilde{\mathcal{V}}$ of the vector space V, denote the change of basis matrix $P = \mathsf{I}_{\tilde{v}v}$. Then, multiply Eq. (9.2) by P^{-1}, that is,

$$P^{-1}\mathsf{T}_{vv}\mathsf{x}_v = \lambda\, P^{-1}\mathsf{x}_v \quad \Leftrightarrow \quad (P^{-1}\mathsf{T}_{vv}P)(P^{-1}\mathsf{x}_v) = \lambda\,(P^{-1}\mathsf{x}_v);$$

using the change of basis formulas $\mathsf{T}_{\tilde{v}\tilde{v}} = P^{-1}\mathsf{T}_{vv}P$ and $\mathsf{x}_{\tilde{v}} = P^{-1}\mathsf{x}_v$ we conclude that

$$\mathsf{T}_{\tilde{v}\tilde{v}}\,\mathsf{x}_{\tilde{v}} = \lambda\,\mathsf{x}_{\tilde{v}}.$$

This establishes the Theorem. □

EXAMPLE 9.1.1: Consider the vector space $\mathbb{R}^2$ with the standard basis $\mathcal{S}$. Find the eigenvalues and eigenvectors of the linear operator $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix in the standard basis given by

$$\mathsf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

SOLUTION: This operator makes a reflection along the line $x_1 = x_2$, that is,

$$\mathsf{T}\mathsf{x} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}.$$

From this definition we see that any non-zero vector proportional to $\mathsf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is left invariant by T, that is,

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Leftrightarrow \quad \mathsf{T}(\mathsf{v}_1) = \mathsf{v}_1.$$

So we conclude that v_1 is an eigenvector of T with eigenvalue $\lambda_1 = 1$. Analogously, one can check that the vector $\mathsf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ satisfies the equation

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = - \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \Leftrightarrow \quad \mathsf{T}(\mathsf{v}_2) = -\mathsf{v}_2.$$

So we conclude that v_2 is an eigenvector of T with eigenvalue $\lambda_2 = -1$. See Fig. 57.

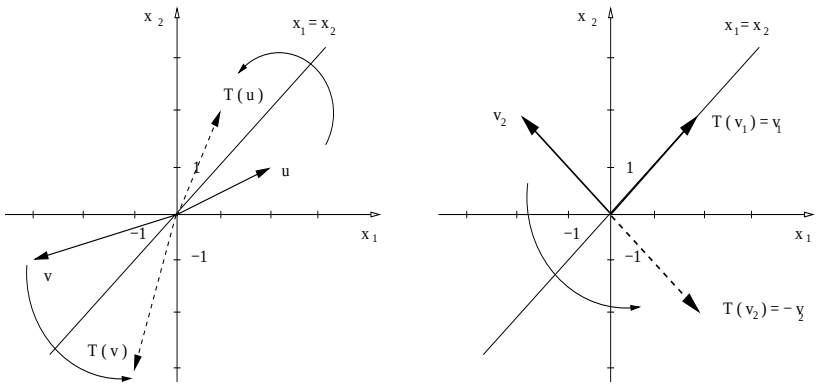

FIGURE 57. On the first picture we sketch the action of matrix T in Example 9.1.1, and on the second picture, and we sketch the eigenvectors v_1 and v_2 with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$, respectively.

In this example the spectrum is $\sigma_T = \{1, -1\}$ and the respective eigenspaces are

$$E_1 = \text{Span}\Big(\Big\{\begin{bmatrix}1\\1\end{bmatrix}\Big\}\Big), \qquad E_{-1} = \text{Span}\Big(\Big\{\begin{bmatrix}-1\\1\end{bmatrix}\Big\}\Big).$$

◁

EXAMPLE 9.1.2: Not every linear operator has eigenvalues and eigenvectors. Consider the vector space $\mathbb{R}^2$ with standard basis and fix $\theta \in (0, \pi)$. The linear operator $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix

$$\mathsf{T} = \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}$$

acts on the plane rotating every vector by an angle θ counterclockwise. Since $\theta \in (0, \pi)$, there is no line on the plane left invariant by the rotation. Therefore, this operator has no eigenvalues and eigenvectors. ◁

EXAMPLE 9.1.3: Consider the vector space $\mathbb{R}^2$ with standard basis. Show that the linear operator $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix $\mathsf{T} = \begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}$ has the eigenvalues and eigenvectors

$$\mathsf{v}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \quad \lambda_1 = 4, \quad \text{and} \quad \mathsf{v}_2 = \begin{bmatrix}1\\-1\end{bmatrix}, \quad \lambda_2 = -2.$$

SOLUTION: We only verify that Eq. (9.1) holds for the vectors and scalars above, since

$$\mathsf{T}\mathsf{v}_1 = \begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}4\\4\end{bmatrix} = 4\begin{bmatrix}1\\1\end{bmatrix} = \lambda_1\mathsf{v}_1 \quad \Rightarrow \quad \mathsf{T}\mathsf{v}_1 = \lambda_1\mathsf{v}_1$$

$$\mathsf{T}\mathsf{v}_2 = \begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}-2\\2\end{bmatrix} = -2\begin{bmatrix}1\\-1\end{bmatrix} = \lambda_2\mathsf{v}_2 \quad \Rightarrow \quad \mathsf{T}\mathsf{v}_2 = \lambda_2\mathsf{v}_2.$$

In this example the spectrum is $\sigma_T = \{4, -2\}$ and the respective eigenspaces are

$$E_4 = \text{Span}\Big(\Big\{\begin{bmatrix}1\\1\end{bmatrix}\Big\}\Big), \qquad E_{-2} = \text{Span}\Big(\Big\{\begin{bmatrix}1\\-1\end{bmatrix}\Big\}\Big).$$

◁

EXAMPLE 9.1.4: Consider the vector space $V = C^\infty(\mathbb{R}, \mathbb{R})$. Show that the vector $\boldsymbol{f}(x) = e^{ax}$, with $a \neq 0$, is an eigenvector with eigenvalue a of the linear operator $\boldsymbol{D} : V \to V$, given by $\boldsymbol{D}(\boldsymbol{f})(x) = \dfrac{d\boldsymbol{f}}{dx}(x)$.

SOLUTION: The proof is straightforward, since

$$\boldsymbol{D}(\boldsymbol{f})(x) = \frac{d\boldsymbol{f}}{dx}(x) = ae^{ax} = a\,\boldsymbol{f}(x) \quad \Rightarrow \quad \boldsymbol{D}(\boldsymbol{f})(x) = a\,\boldsymbol{f}(x).$$

◁

EXAMPLE 9.1.5: Consider again the vector space $V = C^\infty(\mathbb{R}, \mathbb{R})$ and show that the vector $\boldsymbol{f}(x) = \cos(ax)$, with $a \neq 0$, is an eigenvector with eigenvalue $-a^2$ of the linear operator $\boldsymbol{T} : V \to V$, given by $\boldsymbol{T}(\boldsymbol{f})(x) = \dfrac{d^2\boldsymbol{f}}{dx^2}(x)$.

SOLUTION: Again, the proof is straightforward, since

$$\boldsymbol{T}(\boldsymbol{f})(x) = \frac{d^2\boldsymbol{f}}{dx^2}(x) = -a^2\cos(ax) = -a^2\,\boldsymbol{f}(x) \quad \Rightarrow \quad \boldsymbol{T}(\boldsymbol{f})(x) = -a^2\,\boldsymbol{f}(x).$$

◁

We know that an eigenspace of a linear operator is not only a subset of the vector space, it is a subspace. Moreover, it is not any subspace, it is an invariant subspace under the linear operator. Given a vector space V and a linear operator $\boldsymbol{T} \in L(V)$, the subspace $W \subset V$ is *invariant* under $\boldsymbol{T}$ iff holds $\boldsymbol{T}(W) \subset W$.

Theorem 9.1.3. *The eigenspace E_λ of the linear operator $\boldsymbol{T} \in L(V)$ corresponding to the eigenvalue λ is an invariant subspace of the vector space V under the operator $\boldsymbol{T}$.*

Proof of Theorem 9.1.3: We first show that E_λ is a subspace. Indded, pick up any vectors $\boldsymbol{x}$, $\boldsymbol{y} \in E_\lambda$, that is, $\boldsymbol{T}(\boldsymbol{x}) = \lambda\,\boldsymbol{x}$ and $\boldsymbol{T}(\boldsymbol{y}) = \lambda\,\boldsymbol{y}$. Then, for all a, $b \in \mathbb{F}$ holds,

$$\boldsymbol{T}(a\boldsymbol{x} + b\boldsymbol{y}) = a\,\boldsymbol{T}(\boldsymbol{x}) + b\,\boldsymbol{T}(\boldsymbol{y}) = a\lambda\,\boldsymbol{x} + b\lambda\,\boldsymbol{y} = \lambda\,(a\boldsymbol{x} + b\boldsymbol{y}) \quad \Rightarrow \quad (a\boldsymbol{x} + b\boldsymbol{y}) \in E_\lambda.$$

This shows that E_λ is a subspace. Now we need to show that $\boldsymbol{T}(E_\lambda) \subset E_\lambda$. This is straightforward, since for every vector $\boldsymbol{x} \in E_\lambda$ we know that

$$\boldsymbol{T}(\boldsymbol{x}) = \lambda\,\boldsymbol{x} \in E_\lambda,$$

since the set E_λ is a subspace. Therefore, $T(E_\lambda) \subset E_\lambda$. This establishes the Theorem. □

9.1.2. **Characteristic polynomial.** We now address the eigenvalue-eigenvector problem: Given a finite-dimensional vector space V over $\mathbb{F}$ and a linear operator $\boldsymbol{T} \in L(V)$, find a scalar $\lambda \in \mathbb{F}$ and a non-zero vector $\boldsymbol{x} \in V$ solution of $\boldsymbol{T}(\boldsymbol{x}) = \lambda\,\boldsymbol{x}$. This problem is more complicated than solving a linear system of equations $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{b}$, since in our case the source vector is $\boldsymbol{b} = \lambda\boldsymbol{x}$, which is not a given but part of the unknowns. One way to solve the eigenvalue-eigenvector problem is to first solve for the eigenvalues λ and then solve for the eigenvectors $\boldsymbol{x}$.

Theorem 9.1.4. *Let V be a finite-dimensional vector space over $\mathbb{F}$ and let $\boldsymbol{T} \in L(V)$.*

(a) The scalar $\lambda \in \mathbb{F}$ is an eigenvalue of $\boldsymbol{T}$ iff λ is solution of the equation

$$\det(\boldsymbol{T} - \lambda\,\boldsymbol{I}) = 0.$$

(b) Given $\lambda \in \mathbb{F}$ eigenvalue of $\boldsymbol{T}$, the corresponding eigenvectors $\boldsymbol{x} \in V$ are the non-zero solutions of the equation

$$(\boldsymbol{T} - \lambda\,\boldsymbol{I})(\boldsymbol{x}) = \boldsymbol{0}.$$

Remark: The determinant of a linear operator on a finite-dimensional vector space V was introduced in Def. 5.5.6 as the determinant of its associated matrix in any ordered basis of V. This definition is independent of the basis chosen in V, since the operator matrix transforms by a similarity transformation under a change of basis and the determinant is invariant under similarity transformations.

Proof of Theorem 9.1.4:

Part (a): The scalar λ and the vector $\boldsymbol{x}$ are eigenvalue and eigenvector of $\boldsymbol{T}$ iff holds

$$\boldsymbol{T}(\boldsymbol{x}) = \lambda\,\boldsymbol{x} \quad \Leftrightarrow \quad (\boldsymbol{T} - \lambda\,\boldsymbol{I})\,\boldsymbol{x} = \boldsymbol{0} \quad \Leftrightarrow \quad \det(\boldsymbol{T} - \lambda\,\boldsymbol{I}) = 0.$$

Part (b): This is simpler. Since λ is the scalar such that the operator $(\boldsymbol{T} - \lambda\,\boldsymbol{I})$ is not invertible, this means that $N(\boldsymbol{T} - \lambda\,\boldsymbol{I}) \neq \{\boldsymbol{0}\}$, that is, there exists a solution $\boldsymbol{x}$ to the linear equation $(\boldsymbol{T} - \lambda\,\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. It is simple to see that this solution is an eigenvector of $\boldsymbol{T}$. This establishes the Theorem. □

Definition 9.1.5. *Given a finite-dimensional vector space V and a linear operator $\boldsymbol{T} \in L(V)$, the function $p(\lambda) = \det(\boldsymbol{T} - \lambda\,\boldsymbol{I})$ is called the* ***characteristic polynomial*** *of $\boldsymbol{T}$.*

The function p defined above is a polynomial in λ, which can be seen from the definition of determinant of a matrix. The eigenvalues of a linear operator are the roots of its characteristic polynomial.

EXAMPLE 9.1.6: Find the eigenvalues and eigenvectors of the linear operator $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$, given by $\mathsf{T} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

SOLUTION: We start computing the eigenvalues, which are the roots of the characteristic polynomial

$$p(\lambda) = \det(\mathsf{T} - \lambda\,\mathsf{I}) = \det\left(\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \begin{vmatrix} (1-\lambda) & 3 \\ 3 & (1-\lambda) \end{vmatrix},$$

hence $p(\lambda) = (1-\lambda)^2 - 9$. The roots are

$$(\lambda - 1)^2 = 3^2 \quad \Rightarrow \quad \begin{cases} \lambda_1 = 4, \\ \lambda_2 = -2. \end{cases}$$

We now find the eigenvector for the eigenvalue $\lambda_1 = 4$. We solve the system $(\mathsf{T} - 4\,\mathsf{I})\,\mathsf{x} = \mathsf{0}$ performing Gauss operation in the matrix

$$\mathsf{T} - 4\,\mathsf{I} = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = x_2, \\ x_2 \text{ free.} \end{cases}$$

Choosing $x_2 = 1$ we obtain the eigenvector $\mathsf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. In a similar way we find the eigenvector for the eigenvalue $\lambda_2 = -2$. We solve the linear system $(\mathsf{T} + 2\,\mathsf{I})\,\mathsf{x} = \mathsf{0}$ performing Gauss operation in the matrix

$$\mathsf{T} + 2\,\mathsf{I} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = -x_2, \\ x_2 \text{ free.} \end{cases}$$

Choosing $x_2 = -1$ we obtain the eigenvector $\mathsf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. These results λ_1, x_1 and λ_2, x_2 agree with Example 9.1.3. ◁

9.1.3. **Eigenvalue multiplicities.** We now introduce two numbers that give information regarding the size of eigenspaces. The first number determines the maximum possible size of an eigenspace, while the second number characterizes the actual size of an eigenspace.

Definition 9.1.6. *Let $\lambda_i \in \mathbb{F}$, for $i = 1, \cdots, k$ be all the eigenvalues of a linear operator $\boldsymbol{T} \in L(V)$ on a vector space V over $\mathbb{F}$. Express the characteristic polynomial p associated with $\boldsymbol{T}$ as follows,*

$$p(\lambda) = (\lambda - \lambda_1)^{r_1} \cdots (\lambda - \lambda_k)^{r_k} q(\lambda), \quad \text{with} \quad q(\lambda_i) \neq 0,$$

*and denote by $s_i = \dim E_{\lambda_i}$, the dimension of the eigenspaces corresponding to the eigenvalue λ_i. Then, the numbers r_i are called the **algebraic multiplicity** of the eigenvalue λ_i; and the numbers s_i are called the **geometric multiplicity** of the eigenvalue λ_i.*

REMARK: In the case that $\mathbb{F} = \mathbb{C}$, hence the characteristic polynomial is complex-valued, the polynomial $q(\lambda) = 1$. Indeed, when p is an n-degree complex-valued polynomial the Fundamental Theorem of Algebra says that p has n complex roots. Therefore, the characteristic polynomial has the form

$$p(\lambda) = (\lambda - \lambda_1)^{r_1} \cdots (\lambda - \lambda_k)^{r_k},$$

where $r_1 + \cdots + r_k = n$. On the other hand, in the case that $\mathbb{F} = \mathbb{R}$ the characteristic polynomial is real-valued. In this case the polynomial q may have degree greater than zero.

EXAMPLE 9.1.7: For the following matrices find the algebraic and geometric multiplicities of their eigenvalues,

$$\mathsf{A} = \begin{bmatrix} 3 & 2 \\ 0 & 3 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathsf{C} = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

SOLUTION: The eigenvalues of matrix A are the roots of the characteristic polynomial

$$p_a(\lambda) = \begin{vmatrix} (3-\lambda) & 2 \\ 0 & (3-\lambda) \end{vmatrix} = (\lambda - 3)^2 \quad \Rightarrow \quad \lambda_1 = 3, \quad r_1 = 2.$$

So, the eigenvalue $\lambda_1 = 3$ has algebraic multiplicity $r_1 = 2$. To find the geometric multiplicity we need to compute the eigenspace E_{λ_1}, which is the null space of the matrix $(\mathsf{A} - 3\mathsf{I})$, that is,

$$\mathsf{A} - 3\,\mathsf{I}_2 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad x_2 = 0, \quad x_1 \text{ free} \quad \Rightarrow \quad \mathsf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x_1.$$

We have obtained

$$E_{\lambda_1} = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} \right) \quad \Rightarrow \quad s_1 = \dim E_{\lambda_1} = 1.$$

In the case of matrix A the algebraic multiplicity is greater than the geometric multiplicity, $2 = r_1 > s_1 = 1$, since the greatest linearly independent set of eigenvectors for the eigenvalue λ_1 contains only one vector.

The algebraic and geometric multiplicities for matrices B and C is computed in a similar way. These matrices differ only in one single matrix coefficient, the coefficient $(1, 2)$. This difference does not affect their eigenvalues, since we will show that both matrices B and C have the same eigenvalues with the same algebraic multiplicities. However, this difference is enough to change their eigenvectors, since we will show that matrices B and C have different eigenvectors with different geometric multiplicities. We start computing the characteristic polynomial of matrix B,

$$p_b(\lambda) = \begin{vmatrix} (3-\lambda) & 1 & 1 \\ 0 & (3-\lambda) & 2 \\ 0 & 0 & (1-\lambda) \end{vmatrix} = -(\lambda - 3)^2(\lambda - 1),$$

which implies that $\lambda_1 = 3$ has algebraic multiplicity $r_1 = 2$ and $\lambda_2 = 1$ has algebraic multiplicity $r_2 = 1$. To find the geometric multiplicities we need to find the corresponding eigenspaces. We start with $\lambda_1 = 3$,

$$\mathsf{B} - 3\,\mathsf{I}_3 = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & -2 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 \text{ free}, \\ x_2 = 0, \\ x_3 = 0, \end{cases}$$

which implies that

$$E_{\lambda_1} = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\} \right) \quad \Rightarrow \quad s_1 = 1.$$

The geometric multiplicity for the eigenvalue $\lambda_2 = 1$ is computed as follows,

$$\mathsf{B} - \mathsf{I}_3 = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = 0, \\ x_2 = -x_3, \\ x_3 \text{ free}, \end{cases}$$

which implies that

$$E_{\lambda_2} = \text{Span}\Big(\Big\{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}\Big\}\Big) \quad \Rightarrow \quad s_2 = 1.$$

We then conclude $2 = r_1 > s_1 = 1$ and $r_2 = s_2 = 1$.

Finally, we compute the characteristic polynomial of matrix C,

$$p_c(\lambda) = \begin{vmatrix} (3-\lambda) & 0 & 1 \\ 0 & (3-\lambda) & 2 \\ 0 & 0 & (1-\lambda) \end{vmatrix} = -(\lambda-3)^2(\lambda-1)$$

so, we obtain the same eigenvalues we had for matrix B, that is, $\lambda_1 = 3$ has algebraic multiplicity $r_1 = 2$ and $\lambda_2 = 1$ has algebraic multiplicity $r_2 = 1$. The geometric multiplicity of λ_1 is computed as follows,

$$\mathsf{C} - 3\,\mathsf{I}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & -2 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 \text{ free}, \\ x_2 \text{ free}, \\ x_3 = 0, \end{cases}$$

which implies that

$$E_{\lambda_1} = \text{Span}\Big(\Big\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\Big\}\Big) \quad \Rightarrow \quad s_1 = 2.$$

In this case we obtained $2 = r_1 = s_1$. The geometric multiplicity for the eigenvalue $\lambda_2 = 1$ is computed as follows,

$$\mathsf{C} - \mathsf{I}_3 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = -\dfrac{1}{2}\,x_3, \\ x_2 = -x_3, \\ x_3 \text{ free}, \end{cases}$$

which implies that (choosing $x_3 = 2$),

$$E_{\lambda_2} = \text{Span}\Big(\Big\{\begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix}\Big\}\Big) \quad \Rightarrow \quad s_2 = 1.$$

We then conclude $r_1 = s_1 = 2$ and $r_2 = s_2 = 1$.

REMARK: Comparing the results for matrix B and C we see that a change in just one matrix coefficient can change the eigenspaces even in the case where the eigenvalues do not change. In fact, it can be shown that the eigenvalues are continuous functions of the matrix coefficients while the eigenvectors are not continuous functions. This means that a small change in the matrix coefficients produces a small change in the eigenvalues but it might produce a big change in the eigenvectors, just like in this Example. ◁

9.1.4. **Operators with distinct eigenvalues.** The eigenvectors corresponding to distinct eigenvalues form a linearly independent set. In other words, eigenspaces corresponding to different eigenvalues have trivial intersection.

Theorem 9.1.7. *If the eigenvalues $\lambda_1, \cdots, \lambda_k \in \mathbb{F}$, with $k \geqslant 1$, of the linear operator $\boldsymbol{T} \in L(V)$ are all different, then the set of corresponding eigenvectors $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\} \subset V$ is linearly independent.*

Proof of Theorem 9.1.7: If $k = 1$, the Theorem is trivially true, so assume $k \geqslant 2$. Let $c_1, \cdots, c_k \in F$ be scalars such that

$$c_1 \boldsymbol{x}_1 + \cdots + c_k \boldsymbol{x}_k = \boldsymbol{0}. \tag{9.3}$$

Now perform the following two steps: First, apply the operator $\boldsymbol{T}$ to both sides of the equation above. Since $\boldsymbol{T}(\boldsymbol{x}_i) = \lambda_i \boldsymbol{x}_i$, we obtain,

$$c_1 \lambda_1 \, \boldsymbol{x}_1 + \cdots + c_k \lambda_k \, \boldsymbol{x}_k = \boldsymbol{0}. \tag{9.4}$$

Second, multiply Eq. (9.3) by λ_1 and subtract it from Eq.(9.4). The result is

$$c_2(\lambda_2 - \lambda_1) \, \boldsymbol{x}_2 + \cdots + c_k(\lambda_k - \lambda_1) \, \boldsymbol{x}_k = \boldsymbol{0}. \tag{9.5}$$

Notice that all factors $\lambda_i - \lambda_1 \neq 0$ for $i = 2, \cdots, k$. Repeat these two steps: First, apply the operator $\boldsymbol{T}$ on both sides of Eq. (9.5), that is,

$$c_2(\lambda_2 - \lambda_1) \, \lambda_2 \, \boldsymbol{x}_2 + \cdots + c_k(\lambda_k - \lambda_1) \, \lambda_k \, \boldsymbol{x}_k = \boldsymbol{0}; \tag{9.6}$$

second, multiply Eq. (9.5) by λ_2 and subtract it from Eq. (9.6), that is,

$$c_2(\lambda_2 - \lambda_1) \, (\lambda_3 - \lambda_2) \, \boldsymbol{x}_3 + \cdots + c_k(\lambda_k - \lambda_1) \, (\lambda_k - \lambda_2) \, \boldsymbol{x}_k = \boldsymbol{0}. \tag{9.7}$$

Repeat the idea in these two steps until one reaches the equation

$$c_k(\lambda_k - \lambda_1) \cdots (\lambda_k - \lambda_{k-1}) \, \boldsymbol{x}_k = \boldsymbol{0}.$$

Since the eigenvalues λ_i are all different, we conclude that $c_k = 0$. Introducing this information at the very begining we get that

$$c_1 \boldsymbol{x}_1 + \cdots + c_{k-1} \boldsymbol{x}_{k-1} = \boldsymbol{0}.$$

Repeating the whole procedure we conclude that $c_{k-1} = 0$. In this way one shows that all coefficient $c_1 = \cdots = c_k = 0$. Therefore, the set $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\}$ is linearly independent. This establishes the Theorem. □

Further reading. A detailed presentation of eigenvalues and eigenvectors of a matrix can be found in Sections 5.1 and 5.2 in Lay's book [2], while a shorter and deeper summary can be found in Section 7.1 in Meyer's book [3]. Also see Chapter 4 in Hassani's book [1].

9.1.5. **Exercises.**

9.1.1.- Find the spectrum and all eigenspaces of the operators $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathsf{B} : \mathbb{R}^3 \to \mathbb{R}^3$,

$$\mathsf{A} = \begin{bmatrix} -10 & -7 \\ 14 & 11 \end{bmatrix},$$

$$\mathsf{B} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & -1 & 2 \end{bmatrix}.$$

9.1.2.- Show that for $\theta \in (0, \pi)$ the rotation operator $\mathsf{R}(\theta) : \mathbb{R}^2 \to \mathbb{R}^2$ has no eigenvalues, where

$$\mathsf{R}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Consider now matrix above as a linear operator $\mathsf{R}(\theta) : \mathbb{C}^2 \to \mathbb{C}^2$. Show that this linear operator has eigenvalues, and find them.

9.1.3.- Let $\mathsf{A} : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator given by

$$\mathsf{A} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & 1 & h \\ 0 & 0 & 2 \end{bmatrix}.$$

(a) Find all the eigenvalues and their corresponding algebraic multiplicities of the matrix A.

(b) Find the value(s) of $h \in \mathbb{R}$ such that the matrix A above has a two-dimensional eigenspace, and find a basis for this eigenspace.

(c) Set $h = 1$, and find a basis for all the eigenspaces of matrix A above.

9.1.4.- Find all the eigenvalues with their corresponding algebraic multiplicities, and find all the associated eigenspaces of the matrix $\mathsf{A} \in \mathbb{R}^{3,3}$ given by

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

9.1.5.- Let $k \in \mathbb{R}$ and consider the matrix $\mathsf{A} \in \mathbb{R}^{4,4}$ given by

$$\mathsf{A} = \begin{bmatrix} 2 & -2 & 4 & -1 \\ 0 & 3 & k & 0 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(a) Find the eigenvalues of A and their algebraic multiplicity.

(b) Find the number k such that matrix A has an eigenspace E_λ that is two dimensional, and find a basis for this E_λ.

9.1.6.- Comparing the characteristic polynomials for $A \in \mathbb{F}^{n,n}$ and A^T, show that these two matrices have the same eigenvalues.

9.1.7.- Let $\mathsf{A} \in \mathbb{R}^{3,3}$ be an invertible matrix with eigenvalues 2, -1 and 3. Find the eigenvalues of:

(a) A^{-1}.

(b) A^k, for any $k \in \mathbb{N}$.

(c) $\mathsf{A}^2 - \mathsf{A}$.

9.2. Diagonalizable operators

9.2.1. **Eigenvectors and diagonalization.** In this Section we study linear operators on a finite dimensional vector space that have a complete set of eigenvectors. This means that there exists a basis of the vector space formed with eigenvectors of the linear operator. We show that these operators are diagonalizable, that is, the matrix of the operator in the basis of its own eigenvectors is diagonal. We end this Section showing that it is not difficult to define functions of operators in the case that the operator is diagonalizable.

Definition 9.2.1. *A linear operator $\mathbf{T} \in L(V)$ defined on an n-dimensional vector space V has a* ***complete set of eigenvectors*** *iff there exists a linearly independent set formed with n eigenvectors of $\boldsymbol{T}$.*

In other words, a linear operator $\boldsymbol{T}$ has a complete set of eigenvectors iff there exists a basis of V formed by eigenvectors of $\boldsymbol{T}$. Not every linear operator has a complete set of eigenvectors. For example, a linear operator without a complete set of eigenvectors is a rotation on a plane $\boldsymbol{R}(\theta) : \mathbb{R}^2 \to \mathbb{R}^2$ by an angle $\theta \in (0, \pi)$. This particular operator has not eigenvectors at all.

Example 9.2.1: Matrix B in Example 9.1.7 does not have a complete set of eigenvectors. The largest linearly independent set of eigenvectors of matrix B contains only two vectors, one possibility is shown below.

$$\mathsf{B} = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \qquad X = \left\{ \mathsf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathsf{x}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

Matrix C in Example 9.1.7 has a complete set of eigenvectors, indeed,

$$\mathsf{C} = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \qquad X = \left\{ \mathsf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathsf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ \mathsf{x}_3 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} \right\}.$$

◁

We now introduce the notion of a diagonalizable operator.

Definition 9.2.2. *A linear operator $\boldsymbol{T} \in L(V)$ defined on a finite dimensional vector space V is called* ***diagonalizable*** *iff there exists a basis $\tilde{\mathcal{V}}$ of V such that the matrix $\mathsf{T}_{\tilde{v}\tilde{v}}$ is diagonal.*

Recall that a square matrix $\mathsf{D} = [D_{ij}]$ is diagonal iff $D_{ij} = 0$ for $i \neq j$. We denote an $n \times n$ diagonal matrix by $\mathsf{D} = \text{diag}[D_1, \cdots, D_n]$, so we use only one index to label the diagonal elements, $D_{ii} = D_i$. Examples of 3×3 diagonal matrices are given by

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \qquad \mathsf{D} = \begin{bmatrix} D_{11} & 0 & 0 \\ 0 & D_{22} & 0 \\ 0 & 0 & D_{33} \end{bmatrix}.$$

Our first result is to show that these two notions given in Definitions 9.2.1 and 9.2.2 are equivalent.

Theorem 9.2.3. *A linear operator $\boldsymbol{T} \in L(V)$ defined on an n-dimensional vector space is diagonalizable iff the operator $\boldsymbol{T}$ has a complete set of eigenvectors. Furthermore, if $\lambda_i, \boldsymbol{x}_i$, for $i = 1, \cdots, n$, are eigenvalue-eigenvector pairs of $\boldsymbol{T}$, then the matrix of the operator $\boldsymbol{T}$ in the ordered eigenvector basis $\tilde{\mathcal{V}} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ is diagonal with the eigenvalues on the diagonal, that is, $\mathsf{T}_{\tilde{v}\tilde{v}} = \text{diag}[\lambda_1, \cdots, \lambda_n]$.*

Proof of Theorem 9.2.3:
($\Rightarrow$) Since $\boldsymbol{T}$ is diagonalizable, we know that there exists a basis $\tilde{\mathcal{V}} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ such that its matrix is diagonal, that is, $\mathsf{T}_{\tilde{v}\tilde{v}} = \text{diag}[\lambda_1, \cdots, \lambda_n]$. This implies that for $i = 1, \cdots, n$ holds

$$\mathsf{T}_{\tilde{v}\tilde{v}}\, \mathsf{x}_{i\tilde{v}} = \lambda_i\, \mathsf{x}_{i\tilde{v}} \quad \Leftrightarrow \quad \boldsymbol{T}(\boldsymbol{x}_i) = \lambda_i\, \boldsymbol{x}_i.$$

We then conclude that $\boldsymbol{x}_i$ is an eigenvector of the operator $\boldsymbol{T}$ with eigenvalue λ_i. Since $\tilde{\mathcal{V}}$ is a basis of the vector space V, this means that the operator $\boldsymbol{T}$ has a complete set of eigenvectors.

($\Leftarrow$) Since the set of eigenvectors $\tilde{\mathcal{V}} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ of the operator $\boldsymbol{T}$ with corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$ form a basis of V, then the eigenvalue-eigenvector equation for $i = 1, \cdots, n$

$$\boldsymbol{T}(\boldsymbol{x}_i) = \lambda_i\, \boldsymbol{x}_i.$$

imply that the matrix $\mathsf{T}_{\tilde{v}\tilde{v}}$ is diagonal, since

$$\mathsf{T}_{\tilde{v}\tilde{v}} = \left[[\boldsymbol{T}(\boldsymbol{x}_1)]_{\tilde{v}}, \cdots, [\boldsymbol{T}(\boldsymbol{x}_n)]_{\tilde{v}}\right] = \left[\lambda_1\, [\boldsymbol{x}_1]_{\tilde{v}}, \cdots, \lambda_n\, [\boldsymbol{x}_n]_{\tilde{v}}\right] = \left[\lambda_1\, \mathsf{e}_1, \cdots, \lambda_n\, \mathsf{e}_n\right],$$

so we arrive at the equation $\mathsf{T}_{\tilde{v}\tilde{v}} = \text{diag}[\lambda_1, \cdots, \lambda_n]$. This establishes the Theorem. □

EXAMPLE 9.2.2: Show that the linear operator $\boldsymbol{T} \in L(\mathbb{R}^2)$ with matrix $\mathsf{T}_{ss} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ in the standard basis of $\mathcal{S}$ of $\mathbb{R}^2$ is diagonalizable.

SOLUTION: In Example 9.1.6 we obtained that the eigenvalues and eigenvectors of matrix T_{ss} are given by

$$\lambda_1 = 4, \quad \mathsf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \lambda_2 = -2, \quad \mathsf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

We now affirm that the matrix of the linear operator $\boldsymbol{T}$ is diagonal in the ordered basis formed by the eigenvectors above,

$$\tilde{\mathcal{V}} = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right).$$

First notice that the set $\tilde{\mathcal{V}}$ is linearly independent, so it is a basis for $\mathbb{R}^2$. Second, the change of basis matrix $\mathsf{P} = \mathsf{I}_{\tilde{v}s}$ is given by

$$\mathsf{P} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{P}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Third, the result of the change of basis is a diagonal matrix:

$$\mathsf{P}^{-1}\mathsf{T}_{ss}\mathsf{P} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & -2 \end{bmatrix} = \mathsf{T}_{\tilde{v}\tilde{v}}.$$

As stated in Theorem 9.2.3, the diagonal elements in $\mathsf{T}_{\tilde{v}\tilde{v}}$ are precisely the eigenvalues of the operator $\boldsymbol{T}$, in the same order as the eigenvectors in the ordered basis $\tilde{\mathcal{V}}$. ◁

The statement in Definition 9.2.2 can be expressed as a statement between matrices. A matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ is diagonalizable if there exists an invertible matrix $\mathsf{P} \in \mathbb{F}^{n,n}$ and a diagonal matrix $\mathsf{D} \in \mathbb{F}^{n,n}$ such that

$$\mathsf{A} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}.$$

These two notions are equivalent, since A is the matrix of an linear operator $\boldsymbol{T} \in L(V)$ in the standard basis $\mathcal{S}$ of V, that is, $\mathsf{A} = \mathsf{T}_{ss}$. The similarity transformation above can expressed as

$$\mathsf{D} = \mathsf{P}^{-1}\, \mathsf{T}_{ss}\, \mathsf{P}.$$

Denoting $\mathsf{P} = \mathsf{I}_{\tilde{v}s}$ as the change of basis matrix from the standard basis $\mathcal{S}$ to a basis $\tilde{\mathcal{V}}$, we conclude that

$$\mathsf{D} = \mathsf{T}_{\tilde{v}\tilde{v}},$$

That is, the matrix of operator $\boldsymbol{T}$ is diagonal in the basis $\tilde{\mathcal{V}}$. Furthermore, Theorem 9.2.3 can also be expressed in terms of similarity transformations between matrices as follows: *A square matrix has a complete set of eigenvectors iff the matrix is similar to a diagonal matrix*. This active point of view is common in the literature.

EXAMPLE 9.2.3: Show that the linear operator $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix $\mathsf{T} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ in the standard basis of $\mathbb{R}^2$ is diagonalizable. Find a similarity transformation that converts matrix T into a diagonal matrix.

SOLUTION: To find out whether $\boldsymbol{T}$ is diagonalizable or not we need to compute its eigenvectors, so we start with its eigenvalues. The characteristic polynomial is

$$p(\lambda) = \begin{vmatrix} (1-\lambda) & 2 \\ 3 & (6-\lambda) \end{vmatrix} = \lambda(\lambda - 7) = 0 \quad \Rightarrow \quad \begin{cases} \lambda_1 = 0, \\ \lambda_2 = 7. \end{cases}$$

Since the eigenvalues are different, we know that the corresponding eigenvectors form a linearly independent set (by Theorem 9.1.7), and so matrix T has a complete set of eigenvectors. So A is diagonalizable. The corresponding eigenvectors are the non-zero vectors in the null spaces $N(\mathsf{T})$ and $N(\mathsf{T} - 7\,\mathsf{I}_2)$, which are computed as follows:

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad x_1 = -2x_2 \quad \Rightarrow \quad \mathsf{x}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix};$$

$$\begin{bmatrix} -6 & 2 \\ 3 & -1 \end{bmatrix} \to \begin{bmatrix} 3 & -1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad 3x_1 = x_2 \quad \Rightarrow \quad \mathsf{x}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Since the set $\mathcal{V} = \{\mathsf{x}_1, \mathsf{x}_2\}$ is a complete set of eigenvectors of $\boldsymbol{T}$, we conclude that $\boldsymbol{T}$ is diagonalizable. Proposition 9.2.3 says that

$$\mathsf{D} = \mathsf{P}^{-1}\mathsf{T}\,\mathsf{P}, \quad \text{where} \quad \mathsf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix}, \quad \mathsf{P} = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}.$$

We finally verify this the equation above is correct:

$$\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1} = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix} \frac{1}{7} \begin{bmatrix} -3 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \mathsf{T}.$$

◁

9.2.2. **Functions of diagonalizable operators.** We have seen in Section 5.3 that the set of all linear operators $L(V)$ on a vector space V is itself an algebra, since both the linear combination of linear operators in $L(V)$ and the composition of linear operators in $L(V)$ are again linear operators in $L(V)$. We have also seen that this algebra structure on $L(V)$ makes possible to introduce polynomial functions of linear operators. Indeed, given scalars $a_0, \cdots, a_n \in \mathbb{F}$, an operator-valued polynomial of degree n is a function $p : L(V) \to L(V)$,

$$p(\boldsymbol{T}) = a_0 \boldsymbol{I}_V + a_1 \boldsymbol{T} + \cdots + a_n \boldsymbol{T}^n.$$

In this Section we introduce, on the one hand, functions of operators more general than polynomial functions, but on the other hand, we define these functions only on diagonalizable operators. The reason of this restriction is that functions of diagonalizable operators are simple to define. It is possible to define general functions of non-diagonalizable operators, but they are more involved, and will no be studied in these notes.

Before computing a general function of a diagonalizable operator, we start finding simple expressions for the power function and a polynomial function of a diagonalizable operator. These formulas simplify the results we have found in Section 5.3.

Theorem 9.2.4. *If $\boldsymbol{T} \in L(V)$ is a diagonalizable linear operator with matrix $\mathsf{T} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}$ in the standard ordered basis of an n-dimensional vector space V, where D, $\mathsf{P} \in \mathbb{F}^{n,n}$ are a diagonal and an invertible matrix, respectively, then, for every $k \in \mathbb{N}$ holds*

$$\mathsf{T}^k = \mathsf{P}\,\mathsf{D}^k\,\mathsf{P}^{-1}.$$

Proof of Theorem 9.2.4: Consider the case $k = 2$. A simple calculation shows

$$\mathsf{T}^2 = \mathsf{T}\,\mathsf{T} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}\,\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1} = \mathsf{P}\,\mathsf{D}\,\mathsf{D}\,\mathsf{P}^{-1} = \mathsf{P}\,\mathsf{D}^2\,\mathsf{P}^{-1}.$$

Suppose now that the formula holds for $k \in \mathbb{N}$, that is, $\mathsf{T}^k = \mathsf{P}\,\mathsf{D}^k\,\mathsf{P}^{-1}$, and let us show that it also holds for $k+1$. Indeed,

$$\mathsf{T}^{(k+1)} = \mathsf{T}\,\mathsf{T}^k = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}\,\mathsf{P}\,\mathsf{D}^k\,\mathsf{P}^{-1} = \mathsf{P}\,\mathsf{D}\;\mathsf{D}^k\,\mathsf{P}^{-1} = \mathsf{P}\,\mathsf{D}^{(k+1)}\,\mathsf{P}^{-1}.$$

This establishes the Theorem. □

EXAMPLE 9.2.4: Given the matrix $\mathsf{T} \in \mathbb{R}^{3,3}$, compute T^k, where $k \in \mathbb{N}$ and $\mathsf{T} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$.

SOLUTION: From Example 9.2.3 we know that T is diagonalizable, and that

$$\mathsf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix}, \quad \mathsf{P} = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{P}^{-1} = \frac{1}{7}\begin{bmatrix} -3 & 1 \\ 1 & 2 \end{bmatrix}.$$

Therefore,

$$\mathsf{T}^k = \mathsf{P}\,\mathsf{D}^k\,\mathsf{P}^{-1} = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 7^k \end{bmatrix}\frac{1}{7}\begin{bmatrix} -3 & 1 \\ 1 & 2 \end{bmatrix} = 7^{(k-1)}\begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix}\frac{1}{7}\begin{bmatrix} -3 & 1 \\ 1 & 2 \end{bmatrix}.$$

The final result is

$$\mathsf{T}^k = 7^{(k-1)}\,\mathsf{T}.$$

◁

It is simple to generalize Theorem 9.2.4 to polynomial functions.

Theorem 9.2.5. *Assume the hypotheses given in Theorem 9.2.4 and denote the diagonal matrix as $\mathsf{D} = \text{diag}[\lambda_1, \cdots, \lambda_n]$. If $p : \mathbb{F} \to \mathbb{F}$ is a polynomial of degree $k \in \mathbb{N}$, then holds*

$$p(\mathsf{T}) = \mathsf{P}\,p(\mathsf{D})\,\mathsf{P}^{-1}, \quad \text{where} \quad p(\mathsf{D}) = \text{diag}[p(\lambda_1), \cdots, p(\lambda_n)].$$

Proof of Theorem 9.2.5: Given scalars $a_0, \cdots, a_k \in \mathbb{F}$, denote the polynomial p by

$$p(x) = a_0 + a_1 x + \cdots + a_k x^k.$$

Then,

$$\begin{aligned} p(\mathsf{T}) &= a_0\,\mathsf{I}_n + a_1\,\mathsf{T} + \cdots + a_k\,\mathsf{T}^k \\ &= a_0\,\mathsf{P}\,\mathsf{P}^{-1} + a_1\,\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1} + \cdots + a_k\,\mathsf{P}\,\mathsf{D}^k\,\mathsf{P}^{-1} \\ &= \mathsf{P}\,(a_0\,\mathsf{I}_n + a_1\,\mathsf{D} + \cdots + a_k\,\mathsf{D}^k)\,\mathsf{P}^{-1}. \end{aligned}$$

Noticing that

$$p(\mathsf{D}) = a_0\,\mathsf{I}_n + a_1\,\mathsf{D} + \cdots + a_k\,\mathsf{D}^k = \text{diag}\,[p(\lambda_1), \cdots, p(\lambda_n)],$$

we conclude that

$$p(\mathsf{T}) = \mathsf{P}\,p(\mathsf{D})\,\mathsf{P}^{-1}.$$

This establishes the Theorem. □

9.2.3. **The exponential of diagonalizable operators.** Functions that admit a convergent power series expansion can defined on diagonalizable operators in the same way as polynomial functions in Theorem 9.2.5. We consider first an important particular case, the exponential function $f(x) = e^x$. The exponential function is usually defined as the inverse of the natural logarithm function $g(x) = \ln(x)$, which in turns is defined as the area under the graph of the function $h(x) = 1/x$ from 1 to x, that is,

$$\ln(x) = \int_1^x \frac{1}{y}\, dy \qquad x \in (0, \infty).$$

It is not clear how to use this definition of the exponential function on real numbers to extended it to operators. However, one shows that the exponential function on real numbers has several properties, among them that it can be expressed as a convergent infinite power series,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!} + \cdots .$$

It can be proven that defining the exponential function on real numbers as the convergent power series above is equivalent to the definition given earlier as the inverse of the natural logarithm. However, only the power series expression provides the path to generalize the exponential function to a diagonalizable linear operator.

Definition 9.2.6. *The **exponential** of a linear operator $\boldsymbol{T} \in L(V)$ on a finite dimensional vector space, denoted as $e^{\boldsymbol{T}}$, is given by the infinite sum*

$$e^{\boldsymbol{T}} = \sum_{k=0}^{\infty} \frac{\boldsymbol{T}^k}{k!}.$$

The following result says that the infinite sum of operators given in the Definition 9.2.6 converges in the case that the operator is diagonal or the operator is diagonalizable.

Theorem 9.2.7. *If a linear operator $\boldsymbol{T} \in L(V)$ on a finite dimensional vector space is diagonal, that is, there exists an ordered basis such that $\mathsf{T} = \text{diag}[\lambda_1, \cdots, \lambda_n]$, then*

$$e^{\mathsf{T}} = \text{diag}[e^{\lambda_1}, \cdots, e^{\lambda_n}].$$

If a linear operator $\boldsymbol{T} \in L(V)$ on a finite dimensional vector space is diagonalizable, that is, there exists an ordered basis such that $\mathsf{T} = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}$, where $\mathsf{D} = \text{diag}[\lambda_1, \cdots, \lambda_n]$, then

$$e^{\mathsf{T}} = \mathsf{P}\, e^{\mathsf{D}}\, \mathsf{P}^{-1}.$$

Proof of Theorem 9.2.7: Let T be the matrix of the operator $\boldsymbol{T} \in L(V)$ in an ordered basis $\mathcal{V} \subset V$. For every $N \in \mathbb{N}$ introduce the partial sum $S_N(\mathsf{T})$ as follows,

$$S_N(\mathsf{T}) = \sum_{k=0}^{N} \frac{\mathsf{T}^k}{k!},$$

which is a well defined polynomial in T. Assume now that T is diagonalizable, so there exist an invertible matrix P and a diagonal matrix D such that $\mathsf{T} = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}$. Then, a straightforward computation shows that

$$S_N(\mathsf{T}) = \sum_{k=0}^{N} \frac{\mathsf{P}\,\mathsf{D}^k\,\mathsf{P}^{-1}}{k!} = \mathsf{P}\left(\sum_{k=0}^{N} \frac{\mathsf{D}^k}{k!}\right)\mathsf{P}^{-1}. \tag{9.8}$$

In the particular case that T is diagonal, then $\mathsf{P} = \mathsf{I}_n$ and $\mathsf{T} = \text{diag}[\lambda_1, \cdots, \lambda_n]$, so the equation above reduced to

$$S_N(\mathsf{T}) = \sum_{k=0}^{N} \frac{\mathsf{T}^k}{k!} = \text{diag}\Big[\sum_{k=0}^{N} \frac{\lambda_1^k}{k!}, \cdots, \sum_{k=0}^{N} \frac{\lambda_n^k}{k!}\Big].$$

In the expression above we can compute the limit as $N \to \infty$, obtaining

$$e^{\mathsf{T}} = \text{diag}\big[e^{\lambda_1}, \cdots, e^{\lambda_n}\big].$$

This result establishes the first part in Theorem 9.2.7. In the case that T is diagonalizable, we go back to Eq. (9.8), where we now denote $\mathsf{D} = \text{diag}[\lambda_1, \cdots, \lambda_n]$. Since D is a diagonal matrix, we can compute the limit $N \to \infty$ in Eq. (9.8), that is,

$$e^{\mathsf{T}} = S_\infty(\mathsf{T}) = \mathsf{P}\, e^{\mathsf{D}}\, \mathsf{P}^{-1},$$

where we have denoted $e^{\mathsf{D}} = \text{diag}[e^{\lambda_1}, \cdots, e^{\lambda_n}]$. This establishes the Theorem. □

EXAMPLE 9.2.5: For every $t \in \mathbb{R}$ find the value of the exponential function $e^{\mathsf{A}\,t}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}.$$

SOLUTION: From Example 9.2.2 we know that A is diagonalizable, and that $\mathsf{A} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}$, where

$$\mathsf{D} = \begin{bmatrix} 4 & 0 \\ 0 & -2 \end{bmatrix}, \qquad \mathsf{P} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{P}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Therefore,

$$\mathsf{A}\,t = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4t & 0 \\ 0 & -2t \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and Theorem 9.2.7 imply that

$$e^{\mathsf{A}\,t} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{4t} & 0 \\ 0 & e^{-2t} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

◁

The function introduced in Example 9.2.5 above can be seen as an operator-valued function $f : \mathbb{R} \to \mathbb{R}^{2,2}$ given by

$$f(t) = e^{\mathsf{A}\,t}, \qquad \mathsf{A} \in \mathbb{R}^{2,2}.$$

It can be shown that this function is actually differentiable, and that

$$\frac{df}{dt}(t) = \mathsf{A}\, e^{\mathsf{A}\,t}.$$

A more precise statement is the following.

Theorem 9.2.8. *If $\boldsymbol{T} \in L(V)$ is a diagonalizable operator on a finite dimensional vector space over $\mathbb{R}$, then the operator-valued function $\boldsymbol{F} : \mathbb{R} \to L(V)$, defined as $\boldsymbol{F}(x) = e^{\boldsymbol{T}\,x}$ for all $x \in \mathbb{R}$, is differentiable and*

$$\frac{d\boldsymbol{F}}{dx}(x) = \boldsymbol{T}\, e^{\boldsymbol{T}\,x} = e^{\boldsymbol{T}\,x}\, \boldsymbol{T}.$$

Proof of Theorem 9.2.8: Fix an ordered basis $\mathcal{V} \subset V$ and denote by T the matrix of $\boldsymbol{T}$ in that basis. Since T is diagonalizable, then $\mathsf{T} = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}$, with $\mathsf{D} = \text{diag}[\lambda_1, \cdots, \lambda_n]$. Denoting by F the matrix of $\boldsymbol{F}$ in the basis $\mathcal{V}$, it is simple to see that

$$\frac{d\mathsf{F}}{dx}(x) = \frac{d}{dx}\Big(\mathsf{P}\, e^{\mathsf{D}\,x}\, \mathsf{P}^{-1}\Big) = \mathsf{P}\Big(\frac{d}{dx} e^{\mathsf{D}\,x}\Big)\, \mathsf{P}^{-1}.$$

It is not difficult to see that the expression of the far right in equation above is given by

$$\frac{d}{dx} e^{\mathsf{D}\,x} = \text{diag}\left[\frac{d}{dx}\big(e^{\lambda_1 x}\big), \cdots, \frac{d}{dx}\big(e^{\lambda_n x}\big)\right] = \text{diag}\left[\lambda_1 e^{\lambda_1 x}, \cdots, \lambda_n e^{\lambda_n x}\right] = \mathsf{D}\, e^{\mathsf{D}\,x} = e^{\mathsf{D}\,x}\, \mathsf{D},$$

where we used the expression $\mathsf{D} = \text{diag}[\lambda_1, \cdots, \lambda_n]$. Recalling that

$$\mathsf{P}\,\mathsf{D}\, e^{\mathsf{D}\,x}\, \mathsf{P}^{-1} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}\,\mathsf{P}\, e^{\mathsf{D}\,x}\, \mathsf{P}^{-1} = \mathsf{T}\, e^{\mathsf{T}\,x},$$

we conclude that

$$\frac{d}{dx} e^{\mathsf{T}\,x} = \mathsf{T}\, e^{\mathsf{T}\,x} = e^{\mathsf{T}\,x}\, \mathsf{T}.$$

This establishes the Theorem. □

EXAMPLE 9.2.6: Find the derivative of $f(t) = e^{\mathsf{A}\,t}$, where $\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

SOLUTION: From Example 9.2.5 we know that

$$e^{\mathsf{A}\,t} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{4t} & 0 \\ 0 & e^{-2t} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Then, Theorem 9.2.8 implies that

$$\frac{d}{dt} e^{\mathsf{A}\,t} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4e^{4t} & 0 \\ 0 & -2e^{-2t} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

◁

We end this Section presenting a result without proof, that says that given any scalar-valued function with a convergent power series, that function can be extended into an operator-valued function in the case that the operator is diagonalizable.

Theorem 9.2.9. *Let $f : \mathbb{F} \to \mathbb{F}$ be a function given by a power series*

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k,$$

which converges for $|z - z_0| < r$, for some positive real number r. If $L_D(V)$ is the subspace of diagonalizable operators on the finite dimensional vector space V, then the operator-valued function $\boldsymbol{F} : L_D(V) \to L_D(V)$ given by

$$\boldsymbol{F}(\boldsymbol{T}) = \sum_{k=0}^{\infty} c_k (\boldsymbol{T} - z_0\, \boldsymbol{I}_V)^k$$

converges iff every eigenvalue λ_i of $\boldsymbol{T}$ satisfies that $|\lambda_i - z_0| < r$ for all $i = 1, \cdots, n$.

9.2.4. **Exercises.**

9.2.1.- Which of the following matrices cannot be diagonalized?

$$A = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix},$$

$$B = \begin{bmatrix} 2 & 0 \\ 2 & -2 \end{bmatrix},$$

$$C = \begin{bmatrix} 2 & 0 \\ 2 & 2 \end{bmatrix}.$$

9.2.2.- Verify that the matrix

$$A = \begin{bmatrix} 7/5 & 1/5 \\ -1 & 1/2 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 9/10$ and associated eigenvectors

$$x_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad x_2 = \begin{bmatrix} -2 \\ 5 \end{bmatrix}.$$

Use this information to compute

$$\lim_{k \to \infty} A^k.$$

9.2.3.- Given the matrix and vector ,

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}, \; x_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

compute the function $x : \mathbb{R} \to \mathbb{R}^2$

$$x(t) = e^{A\,t} x_0.$$

Verify that this function is solution of the differential equation

$$\frac{d}{dt} x(t) = A\, x(t)$$

and satisfies that $x(t = 0) = x_0$.

9.2.4.- Let $A \in \mathbb{R}^{3,3}$ be a matrix with eigenvalues 2, -1 and 3. Find the determinant of A.

9.2.5.- Let $A \in \mathbb{R}^{4,4}$ be a matrix that can be decomposed as $A = P\,D\,P^{-1}$, with matrix P an invertible matrix and the matrix

$$D = \text{diag}(2, \frac{1}{4}, 2, 3).$$

Knowing only this information about the matrix A, is it possible to compute the $\det(A)$? If your answer is ***no***, explain why not; if your answer is ***yes***, compute $\det(A)$ and show your work.

9.2.6.- Let $A \in \mathbb{R}^{4,4}$ be a matrix that can be decomposed as $A = P\,D\,P^{-1}$, with matrix P an invertible matrix and the matrix

$$D = \text{diag}(2, 0, 2, 5).$$

Knowing only this information about the matrix A, is it possible to whether A invertible? Is it possible to know $\text{tr}\,(A)$? If your answer is ***no***, explain why not; if your answer is ***yes***, compute $\text{tr}\,(A)$ and show your work.

9.3. Differential equations

Eigenvalues and eigenvectors of a matrix are useful to find solutions to systems of differential equations. In this Section we first recall what is a system of first order, linear, homogeneous, differential equations with constant coefficients. Then we use the eigenvalues and eigenvectors of the coefficient matrix to obtain solutions to such differential equations.

In order to introduce a linear, first order system of differential equations we need some notation. Let $\mathsf{A} : \mathbb{R} \to \mathbb{R}^{n,n}$ be a real matrix-valued function, $\mathsf{x}, \mathsf{b} : \mathbb{R} \to \mathbb{R}^n$ be real vector-valued functions, with values $\mathsf{A}(t)$, $\mathsf{x}(t)$, and $\mathsf{b}(t)$ given by

$$\mathsf{A}(t) = \begin{bmatrix} A_{11}(t) & \cdots & A_{1n}(t) \\ \vdots & & \vdots \\ A_{n1}(t) & \cdots & A_{nn}(t) \end{bmatrix}, \qquad \mathsf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad \mathsf{b}(t) = \begin{bmatrix} b_1(t) \\ \vdots \\ b_n(t) \end{bmatrix}.$$

So, $\mathsf{A}(t)$ is an $n \times n$ matrix for each value of $t \in \mathbb{R}$. An example in the case $n = 2$ is the matrix-valued function

$$\mathsf{A}(t) = \begin{bmatrix} \cos(2\pi t) & -\sin(2\pi t) \\ \sin(2\pi t) & \cos(2\pi t) \end{bmatrix}.$$

The values of this function are rotation matrices on $\mathbb{R}^2$, counterclockwise by an angle $2\pi t$. So the bigger the parameter t the bigger the is rotation. Derivatives of matrix- and vector-valued functions are computed component-wise, and we use the notation $\dot{} = \frac{d}{dt}$; for example

$$\frac{d\mathsf{x}}{dt}(t) = \begin{bmatrix} \dfrac{dx_1}{dt}(t) \\ \vdots \\ \dfrac{dx_n}{dt}(t) \end{bmatrix} \quad \text{is denoted as} \quad \dot{\mathsf{x}}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix}.$$

We are now ready to introduce the main definitions.

Definition 9.3.1. *A* **system of first order linear differential equations** *on n unknowns, with $n \geqslant 1$, is the following: Given a real matrix-valued function $\mathsf{A} : \mathbb{R} \to \mathbb{R}^{n,n}$, and a real vector-valued function $\mathsf{b} : \mathbb{R} \to \mathbb{R}^n$, find a vector-valued function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^n$ solution of*

$$\dot{\mathsf{x}}(t) = \mathsf{A}(t)\,\mathsf{x}(t) + \mathsf{b}(t). \tag{9.9}$$

The system in (9.9) is called **homogeneous** *iff $\mathsf{b}(t) = \mathsf{0}$ for all $t \in \mathbb{R}$. The system in (9.9) is called of* **constant coefficients** *iff the matrix- and vector-valued functions are constant, that is, $\mathsf{A}(t) = \mathsf{A}_0$ and $\mathsf{b}(t) = \mathsf{b}_0$ for all $t \in \mathbb{R}$.*

The differential equation in (9.9) is called first order because it contains only first derivatives of the the unknown vector-valued function x; it is called linear because the unknown x appears linearly in the equation. In this Section we are interested in finding solutions to an initial value problem involving a constant coefficient, homogeneous differential system.

Definition 9.3.2. *An* **initial value problem** *(IVP) for an homogeneous constant coefficients linear differential equation is the following: Given a matrix $\mathsf{A} \in \mathbb{R}^{n,n}$ and a vector $\mathsf{x}_0 \in \mathbb{R}^n$, find a vector-valued function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^n$ solution of the differential equation and initial condition*

$$\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0.$$

In this Section we only consider the case where the coefficient matrix $\mathsf{A} \in \mathbb{R}^{n,n}$ has a complete set of eigenvectors, that is, matrix A is diagonalizable. In this case it is possible to find all solutions of the initial value problem for an homogeneous and constant coefficient differential equation. These solutions are linear combination of vectors proportional to the

eigenvectors of matrix A, where the scalars involved in the linear combination depend on the eigenvalues of matrix A. The explicit form of the solution depends on these eigenvalues and can be classified in three groups: Non-repeated real eigenvalues, non-repeated complex eigenvalues, and repeated eigenvalues. We consider in these notes only the first two cases of non-repeated eigenvalues. In this case, the main result is the following.

Theorem 9.3.3. *Assume that matrix $\mathsf{A} \in \mathbb{R}^{n,n}$ has a complete set of eigenvectors denoted as $\mathcal{V} = \{\mathsf{v}_1, \cdots, \mathsf{v}_n\}$ with corresponding eigenvalues $\{\lambda_1, \cdots, \lambda_n\}$, all different, and fix $\mathsf{x}_0 \in \mathbb{R}^n$. Then the initial value problem*

$$\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0, \tag{9.10}$$

has a unique solution given by

$$\mathsf{x}(t) = e^{\mathsf{A}t}\,\mathsf{x}_0, \tag{9.11}$$

where

$$e^{\mathsf{A}t} = \mathsf{P}\,e^{\mathsf{D}t}\,\mathsf{P}^{-1}, \qquad \mathsf{P} = \left[\mathsf{v}_1, \cdots, \mathsf{v}_n\right], \qquad e^{\mathsf{D}t} = diag\left[e^{\lambda_1 t}, \cdots, e^{\lambda_n t}\right].$$

The solution given in Eq. (9.11) is often written in an equivalent way, as follows:

$$\mathsf{x}(t) = \mathsf{P}\,e^{\mathsf{D}t}\,\mathsf{P}^{-1}\,\mathsf{x}_0 = \left(\mathsf{P}\,e^{\mathsf{D}t}\right)\left(\mathsf{P}^{-1}\,\mathsf{x}_0\right),$$

then, introducing the notation

$$\mathsf{P}\,e^{\mathsf{D}t} = \left[\mathsf{v}_1\,e^{\lambda_1 t}, \cdots, \mathsf{v}_n\,e^{\lambda_n t}\right], \qquad \mathsf{c} = \mathsf{P}^{-1}\,\mathsf{x}_0, \qquad \mathsf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix},$$

we write the solution $\mathsf{x}(t)$ in the form

$$\mathsf{x}(t) = c_1\,\mathsf{v}_1\,e^{\lambda_1 t} + \cdots + c_n\,\mathsf{v}_n\,e^{\lambda_n t}, \qquad \mathsf{P}\,\mathsf{c} = \mathsf{x}_0.$$

This latter notation is common in the literature on ordinary differential equations. The solution $\mathsf{x}(t)$ is expressed as a linear combination of the eigenvectors v_i of the coefficient matrix A, where the components are functions of the variable t given by $c_i e^{\lambda_i t}$, for $i = 1, \cdots, n$. So the eigenvalues and eigenvectors of matrix A are the crucial information to find the solution $\mathsf{x}(t)$ of the initial value problem in Eq. (9.3.2), as can be seen from the following calculation: Consider the function

$$\mathsf{y}_i(t) = e^{\lambda_i t}\mathsf{v}_i, \qquad i = 1, \cdots, n.$$

This function is solution of the differential equation above, since

$$\dot{\mathsf{y}}_i(t) = \frac{d}{dt}\left(e^{\lambda_i t}\right)\mathsf{v}_i = \lambda_i e^{\lambda_i t}\,\mathsf{v}_i = e^{\lambda_i t}\left(\lambda_i \mathsf{v}_i\right) = e^{\lambda_i t}\mathsf{A}\,\mathsf{v}_i = \mathsf{A}\left(e^{\lambda_i t}\mathsf{v}_i\right) = \mathsf{A}\,\mathsf{y}_i(t),$$

hence, $\dot{\mathsf{y}}_i(t) = \mathsf{A}\,\mathsf{y}_i(t)$. This calculation is the essential part in the proof of Theorem 9.3.3.

Proof of Theorem 9.3.3: Since matrix A has a complete set of eigenvectors $\mathcal{V}$, then for every value of $t \in \mathbb{R}$ there exist t dependent scalars $c_1(t), \cdots, c_n(t)$ such that

$$\mathsf{x}(t) = c_1(t)\,\mathsf{v}_1 + \cdots + c_n(t)\,\mathsf{v}_n. \tag{9.12}$$

The t-derivative of the expression above is

$$\dot{\mathsf{x}}(t) = \dot{c}_1(t)\,\mathsf{v}_1 + \cdots + \dot{c}_n(t)\,\mathsf{v}_n.$$

Recalling that $\mathsf{A}\,\mathsf{v}_i = \lambda_i\,\mathsf{v}_i$, the action of matrix A in Eq. (9.12) is

$$\mathsf{A}\,\mathsf{x}(t) = c_1(t)\,\lambda_1\,\mathsf{v}_1 + \cdots + c_n(t)\,\lambda_n\,\mathsf{v}_n.$$

The vector $\mathsf{x}(t)$ is solution of the differential equation in (9.3.2) iff $\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t)$, that is,

$$\dot{c}_1(t)\,\mathsf{v}_1 + \cdots + \dot{c}_n(t)\,\mathsf{v}_n. = c_1(t)\,\lambda_1\,\mathsf{v}_1 + \cdots + c_n(t)\,\lambda_n\,\mathsf{v}_n,$$

which is equivalent to

$$[\dot{c}_1(t) - \lambda_1\, c_1(t)]\, \mathsf{v}_1 + \cdots + [\dot{c}_n(t) - \lambda_n\, c_n(t)]\, \mathsf{v}_n = \mathsf{0}.$$

Since the set $\mathcal{V}$ is a basis of $\mathbb{R}^n$, each term above must vanish, that is, for all $i = 1, \cdots, n$ holds

$$\dot{c}_i(t) = \lambda_i\, c_i(t) \quad \Rightarrow \quad c_i(t) = c_i e^{\lambda_i t}.$$

So we have obtained the general solution

$$\mathsf{x}(t) = c_1 e^{\lambda_1 t}\, \mathsf{v}_1 + \cdots + c_n e^{\lambda_n t}\, \mathsf{v}_n.$$

The initial condition $\mathsf{x}(0) = \mathsf{x}_0$ fixes a unique set of constants c_i as follows

$$\mathsf{x}(0) = c_1\, \mathsf{v}_1 + \cdots + c_n\, \mathsf{v}_n = \mathsf{x}_0,$$

since the set $\mathcal{V}$ is a basis of $\mathbb{R}^n$. This expression of the solution can be rewritten as follows: Using the matrix notation $\mathsf{P} = \big[\mathsf{v}_1, \cdots, \mathsf{v}_n\big]$, we see that the vector c satisfies the equation $\mathsf{P}\,\mathsf{c} = \mathsf{x}_0$. Also notice that

$$\mathsf{P}\, e^{\mathsf{D}t} = \big[\mathsf{v}_1\, e^{\lambda_1 t}, \cdots, \mathsf{v}_n\, e^{\lambda_n t}\big],$$

therefore, the solution $\mathsf{x}(t)$ can be written as

$$\mathsf{x}(t) = \big(\mathsf{P}\, e^{\mathsf{D}t}\big)\, \big(\mathsf{P}^{-1}\mathsf{x}_0\big) = \big(\mathsf{P}\, e^{\mathsf{D}t}\, \mathsf{P}^{-1}\big)\, \mathsf{x}_0.$$

Since $e^{\mathsf{A}t} = \mathsf{P}\, e^{\mathsf{D}t}\, \mathsf{P}^{-1}$, we conclude that $\mathsf{x}(t) = e^{\mathsf{A}t}\, \mathsf{x}_0$. This establishes the Theorem. □

EXAMPLE 9.3.1: **(Non-repeated, real eigenvalues)** Given the matrix $\mathsf{A} \in \mathbb{R}^{2,2}$ and vector $\mathsf{x}_0 \in \mathbb{R}^2$ below, find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solution of the initial value problem

$$\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}, \qquad \mathsf{x}_0 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

SOLUTION: Recall that matrix A has a complete set of eigenvectors, with

$$\mathcal{V} = \left\{ \mathsf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ \mathsf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \right\}, \qquad \{\lambda_1 = 4,\ \lambda_2 = -2\}.$$

Then, Theorem 9.3.3 says that the general solution of the differential equation above is

$$\mathsf{x}(t) = c_1\, e^{\lambda_1 t}\, \mathsf{v}_1 + c_2\, e^{\lambda_2 t}\, \mathsf{v}_2 \quad \Leftrightarrow \quad \mathsf{x}(t) = c_1\, e^{4t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2\, e^{-2t} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The constants c_1 and c_2 are obtained from the initial data x_0 as follows

$$\mathsf{x}(0) = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \mathsf{x}_0 = \begin{bmatrix} 6 \\ 4 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

The solution of this linear system is

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}.$$

Therefore, the solution $\mathsf{x}(t)$ of the initial value problem above is

$$\mathsf{x}(t) = 5\, e^{4t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} - e^{-2t} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

◁

9.3.1. **Non-repeated real eigenvalues.** We present a qualitative description of the solutions to Eq. (9.10) in the particular case of 2×2 linear ordinary differential systems with matrix A having two real and different eigenvalues. The main tool will be the sketch of **phase diagrams**, also called phase portraits. The solution at a particular value t is given by a vector

$$\mathsf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

so it can be represented by a point on a plane, while the solution function for all $t \in \mathbb{R}$ corresponds to a curve on that plane. In the case that the solution vector $\mathsf{x}(t)$ represents a position function of a particle moving on the plane at the time t, the curve given in the phase diagram is the trajectory of the particle. Arrows are added to this trajectory to indicate the motion of the particle as time increases.

Since the eigenvalues of the coefficient matrix A are different, Theorem 9.3.3 says that there always exist two linearly independent eigenvectors v_1 and v_2 associated with the eigenvalues λ_1 and λ_2, respectively. The general solution to the Eq. (9.10) is then given by

$$\mathsf{x}(t) = c_1\, \mathsf{v}_1\, e^{\lambda_1 t} + c_2\, \mathsf{v}_2\, e^{\lambda_2 t}.$$

A phase diagram contains several curves associated with several solutions, that correspond to different values of the free constants c_1 and c_2. In the case that the eigenvalues are non-zero, the phase diagrams can be classified into three main classes according to the relative signs of the eigenvalues $\lambda_1 \neq \lambda_2$ of the coefficient matrix A, as follows:

(i) $0 < \lambda_2 < \lambda_1$, that is, both eigenvalues positive;
(ii) $\lambda_2 < 0 < \lambda_1$, that is, one eigenvalue negative and the other positive;
(iii) $\lambda_2 < \lambda_1 < 0$, that is, both eigenvalues negative.

The study of the cases where one of the eigenvalues vanishes is simpler and is left as an exercise. We now find the phase diagrams for three examples, one for each of the classes presented above. These examples summarize the behavior of the solutions to 2×2 linear differential systems with coefficient matrix having two real, different and non-zero eigenvalues $\lambda_2 < \lambda_1$. The phase diagrams can be sketched following these steps: First, plot the eigenvectors v_2 and v_1 corresponding to the eigenvalues λ_2 and λ_1, respectively. Second, draw the whole lines parallel to these vectors and passing through the origin. These straight lines correspond to solutions with one of the coefficients c_1 or c_2 vanishing. Arrows on these lines indicate how the solution changes as the variable t grows. If t is interpreted as time, the arrows indicate how the solution changes into the future. The arrows point towards the origin if the corresponding eigenvalue λ is negative, and they point away form the origin if the eigenvalue is positive. Finally, find the non-straight curves correspond to solutions with both coefficient c_1 and c_2 non-zero. Again, arrows on these curves indicate the how the solution moves into the future.

EXAMPLE 9.3.2: **(Case $0 < \lambda_2 < \lambda_1$.)** Sketch the phase diagram of the solutions to the differential equation

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x}, \qquad \mathsf{A} = \frac{1}{4}\begin{bmatrix} 11 & 3 \\ 1 & 9 \end{bmatrix}. \tag{9.13}$$

SOLUTION: The characteristic equation for matrix A is given by

$$\det(\mathsf{A} - \lambda\, \mathsf{I}_2) = \lambda^2 - 5\lambda + 6 = 0 \quad \Rightarrow \quad \begin{cases} \lambda_1 = 3, \\ \lambda_2 = 2. \end{cases}$$

One can show that the corresponding eigenvectors are given by

$$\mathsf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \qquad \mathsf{v}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}.$$

So the general solution to the differential equation above is given by

$$\mathsf{x}(t) = c_1\,\mathsf{v}_1 e^{\lambda_1 t} + c_2\,\mathsf{v}_2 e^{\lambda_2 t} \quad \Leftrightarrow \quad \mathsf{x}(t) = c_1 \begin{bmatrix} 3 \\ 1 \end{bmatrix} e^{3t} + c_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} e^{2t}.$$

In Fig. 58 we have sketched four curves, each representing a solution $\mathsf{x}(t)$ corresponding to a particular choice of the constants c_1 and c_2. These curves actually represent eight different solutions, for eight different choices of the constants c_1 and c_2, as is described below. The arrows on these curves represent the change in the solution as the variable t grows. Since both eigenvalues are positive, the length of the solution vector always increases as t grows. The straight lines correspond to the following four solutions:

$c_1 = 1,$	$c_2 = 0,$	Line on the first quadrant, starting at the origin, parallel to v_1;
$c_1 = 0,$	$c_2 = 1,$	Line on the second quadrant, starting at the origin, parallel to v_2;
$c_1 = -1,$	$c_2 = 0,$	Line on the third quadrant, starting at the origin, parallel to $-\mathsf{v}_1$;
$c_1 = 0,$	$c_2 = -1,$	Line on the fourth quadrant, starting at the origin, parallel to $-\mathsf{v}_2$.

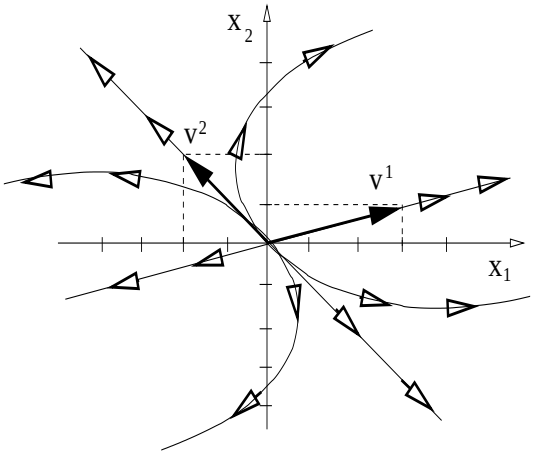

FIGURE 58. The graph of several solutions to Eq. (9.13) corresponding to the case $0 < \lambda_2 < \lambda_1$, for different values of the constants c_1 and c_2. The trivial solution $\mathsf{x} = \mathsf{0}$ is called an unstable point.

Finally, the curved lines on each quadrant start at the origin, and they correspond to the following choices of the constants:

$c_1 > 0,$	$c_2 > 0,$	Line starting on the second to the first quadrant;
$c_1 < 0,$	$c_2 > 0,$	Line starting on the second to the third quadrant;
$c_1 < 0,$	$c_2 < 0,$	Line starting on the fourth to the third quadrant,
$c_1 > 0,$	$c_2 < 0,$	Line starting on the fourth to the first quadrant.

◁

EXAMPLE 9.3.3: **(Case $\lambda_2 < 0 < \lambda_1$.)** Sketch the phase diagram of the solutions to the differential equation

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x}, \qquad \mathsf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}. \tag{9.14}$$

SOLUTION: We known from the calculations performed in Example 9.3.1 that the general solution to the differential equation above is given by

$$\mathsf{x}(t) = c_1\,\mathsf{v}_1 e^{\lambda_1 t} + c_2\,\mathsf{v}_2 e^{\lambda_2 t} \quad \Leftrightarrow \quad \mathsf{x}(t) = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{4t} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t},$$

where we have introduced the eigenvalues and eigenvectors

$$\lambda_1 = 4, \quad \mathsf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \lambda_2 = -2, \quad \mathsf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

In Fig. 59 we have sketched four curves, each representing a solution $\boldsymbol{x}(t)$ corresponding to a particular choice of the constants c_1 and c_2. These curves actually represent eight different solutions, for eight different choices of the constants c_1 and c_2, as is described below. The arrows on these curves represent the change in the solution as the variable t grows. The part of the solution with positive eigenvalue increases exponentially when t grows, while the part of the solution with negative eigenvalue decreases exponentially when t grows. The straight lines correspond to the following four solutions:

$c_1 = 1,$	$c_2 = 0,$	Line on the first quadrant, starting at the origin, parallel to v_1;
$c_1 = 0,$	$c_2 = 1,$	Line on the second quadrant, ending at the origin, parallel to v_2;
$c_1 = -1,$	$c_2 = 0,$	Line on the third quadrant, starting at the origin, parallel to $-\mathsf{v}_1$;
$c_1 = 0,$	$c_2 = -1,$	Line on the fourth quadrant, ending at the origin, parallel to $-\mathsf{v}_2$.

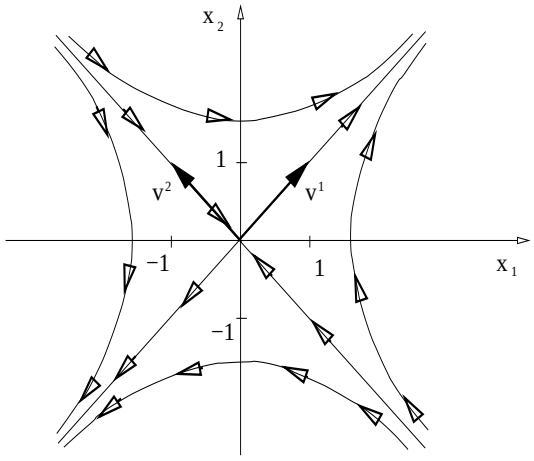

FIGURE 59. The graph of several solutions to Eq. (9.14) corresponding to the case $\lambda_2 < 0 < \lambda_1$, for different values of the constants c_1 and c_2. The trivial solution $\mathsf{x} = \mathsf{0}$ is called a saddle point.

Finally, the curved lines on each quadrant correspond to the following choices of the constants:

$c_1 > 0,$	$c_2 > 0,$	Line from the second to the first quadrant,
$c_1 < 0,$	$c_2 > 0,$	Line from the second to the third quadrant,
$c_1 < 0,$	$c_2 < 0,$	Line from the fourth to the third quadrant,
$c_1 > 0,$	$c_2 < 0,$	Line from the fourth to the first quadrant.

◁

EXAMPLE 9.3.4: **(Case $\lambda_2 < \lambda_1 < 0$.)** Sketch the phase diagram of the solutions to the differential equation

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x}, \qquad \mathsf{A} = \frac{1}{4}\begin{bmatrix} -9 & 3 \\ 1 & -11 \end{bmatrix}. \tag{9.15}$$

SOLUTION: The characteristic equation for this matrix A is given by

$$\det(\mathsf{A} - \lambda\,\mathsf{I}) = \lambda^2 + 5\lambda + 6 = 0 \quad \Rightarrow \quad \begin{cases} \lambda_1 = -2, \\ \lambda_2 = -3. \end{cases}$$

One can show that the corresponding eigenvectors are given by

$$\mathsf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \qquad \mathsf{v}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}.$$

So the general solution to the differential equation above is given by

$$\mathsf{x}(t) = c_1\,\mathsf{v}_1 e^{\lambda_1 t} + c_2\,\mathsf{v}_2 e^{\lambda_2 t} \quad \Leftrightarrow \quad \mathsf{x}(t) = c_1 \begin{bmatrix} 3 \\ 1 \end{bmatrix} e^{-2t} + c_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} e^{-3t}.$$

In Fig. 60 we have sketched four curves, each representing a solution $\boldsymbol{x}(t)$ corresponding to a particular choice of the constants c_1 and c_2. These curves actually represent eight different solutions, for eight different choices of the constants c_1 and c_2, as is described below. The arrows on these curves represent the change in the solution as the variable t grows. Since both eigenvalues are negative, the length of the solution vector always decreases as t grows and the solution vector always approaches zero. The straight lines correspond to the following four solutions:

$c_1 = 1,$	$c_2 = 0,$	Line on the first quadrant, ending at the origin, parallel to v_1;
$c_1 = 0,$	$c_2 = 1,$	Line on the second quadrant, ending at the origin, parallel to v_2;
$c_1 = -1,$	$c_2 = 0,$	Line on the third quadrant, ending at the origin, parallel to $-\mathsf{v}_1$;
$c_1 = 0,$	$c_2 = -1,$	Line on the fourth quadrant, ending at the origin, parallel to $-\mathsf{v}_2$.

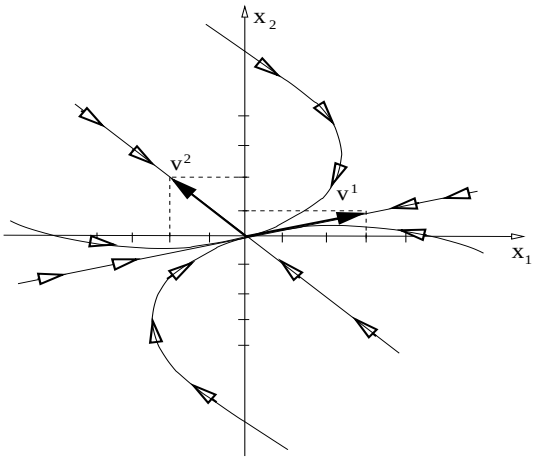

FIGURE 60. The graph of several solutions to Eq. (9.15) corresponding to the case $\lambda_2 < \lambda_1 < 0$, for different values of the constants c_1 and c_2. The trivial solution $\mathsf{x} = \mathsf{0}$ is called a stable point.

Finally, the curved lines on each quadrant start at the origin, and they correspond to the following choices of the constants:

$c_1 > 0,$	$c_2 > 0,$	Line entering the first from the second quadrant;
$c_1 < 0,$	$c_2 > 0,$	Line entering the third from the second quadrant;
$c_1 < 0,$	$c_2 < 0,$	Line entering the third from the fourth quadrant,
$c_1 > 0,$	$c_2 < 0,$	Line entering the first from the fourth quadrant.

◁

9.3.2. **Non-repeated complex eigenvalues.** The complex eigenvalues of a real valued matrix $\mathsf{A} \in \mathbb{R}^{n,n}$ are always complex conjugate pairs, as it is shown below.

Lemma 9.3.4. (Conjugate pairs) *If a real valued matrix $\mathsf{A} \in \mathbb{R}^{n,}$ has a complex eigenvalue λ with eigenvector v, then $\overline{\lambda}$ and $\overline{\mathsf{v}}$ are also an eigenvalue and eigenvector of matrix A.*

Proof of Lemma 9.3.4: Complex conjugate the eigenvalue-eigenvector equation for λ and v and recalling that $\mathsf{A} = \overline{\mathsf{A}}$, we obtain

$$\mathsf{A}\,\mathsf{v} = \lambda\,\mathsf{v} \quad \Leftrightarrow \quad \mathsf{A}\,\overline{\mathsf{v}} = \overline{\lambda}\,\overline{\mathsf{v}}.$$

□

Since the complex eigenvalues of a matrix with real coefficients are always complex conjugate pairs, there is an even number of complex eigenvalues. Denoting the eigenvalue pair by $\lambda_\pm$ and the corresponding eigenvector pair by $\mathsf{v}_\pm$, it holds that $\lambda_+ = \overline{\lambda}_-$ and $\mathsf{v}_+ = \overline{\mathsf{v}}_-$. Hence, an eigenvalue and eigenvector pairs have the form

$$\lambda_\pm = \alpha \pm i\beta, \qquad \mathsf{v}_\pm = \mathsf{a} \pm i\mathsf{b}, \tag{9.16}$$

where α, $\beta \in \mathbb{R}$ and a, $\mathsf{b} \in \mathbb{R}^n$. It is simple to obtain two linearly independent solutions to the differential equation in Eq. (9.10) in the case that matrix A has a complex conjugate pair of eigenvalues and eigenvectors. These solutions can be expressed both as complex-valued or as real-valued functions.

Theorem 9.3.5. (Conjugate pairs) *Let $\lambda_\pm = \alpha \pm i\beta$ be eigenvalues of a matrix $\mathsf{A} \in \mathbb{R}^{n,n}$ with respective eigenvectors $\mathsf{v}_\pm = \mathsf{a} \pm i\mathsf{b}$, where α, $\beta \in \mathbb{R}$, while a, $\mathsf{b} \in \mathbb{R}^n$, and $n \geqslant 2$. Then a linearly independent set of complex valued solutions to the differential equation in (9.10) is formed by the functions*

$$\mathsf{x}_+ = \mathsf{v}_+\, e^{\lambda_+ t}, \qquad \mathsf{x}_- = \mathsf{v}_-\, e^{\lambda_- t}, \tag{9.17}$$

while a linearly independent set of real valued solutions to Eq. (9.10) is given by the functions

$$\mathsf{x}_1 = \big[\mathsf{a}\,\cos(\beta t) - \mathsf{b}\,\sin(\beta t)\big]\, e^{\alpha t}, \qquad \mathsf{x}_2 = \big[\mathsf{a}\,\sin(\beta t) + \mathsf{b}\,\cos(\beta t)\big]\, e^{\alpha t}. \tag{9.18}$$

Proof of Theorem 9.3.5: We know from Theorem 9.3.3 that two linearly independent solutions to Eq. (9.10) are given by Eq. (9.17). The new information in Theorem 9.3.5 above is the real-valued solutions in Eq. (9.18). They can be obtained from Eq. (9.17) as follows:

$$\begin{aligned}
\mathsf{x}_\pm &= (\mathsf{a} \pm i\mathsf{b})\, e^{(\alpha \pm i\beta)t} \\
&= e^{\alpha t}(\mathsf{a} \pm i\mathsf{b})\, e^{\pm i\beta t} \\
&= e^{\alpha t}(\mathsf{a} \pm i\mathsf{b})\,\big[\cos(\beta t) \pm i\sin(\beta t)\big] \\
&= e^{\alpha t}\big[\mathsf{a}\,\cos(\beta t) - \mathsf{b}\,\sin(\beta t)\big] \pm i e^{\alpha t}\big[\mathsf{a}\,\sin(\beta t) + \mathsf{b}\,\cos(\beta t)\big].
\end{aligned}$$

Since the differential equation in (9.10) is linear, the functions below are also solutions,

$$\mathsf{x}_1 = \frac{1}{2}\big(\mathsf{x}_+ + \mathsf{x}_-\big) = e^{\alpha t}\big[\mathsf{a}\,\cos(\beta t) - \mathsf{b}\,\sin(\beta t)\big],$$
$$\mathsf{x}_2 = \frac{1}{2i}\big(\mathsf{x}_+ - \mathsf{x}_-\big) = e^{\alpha t}\big[\mathsf{a}\,\sin(\beta t) + \mathsf{b}\,\cos(\beta t)\big].$$

This establishes the Theorem. □

EXAMPLE 9.3.5: Find a real-valued set of fundamental solutions to the differential equation

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x}, \qquad \mathsf{A} = \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}, \tag{9.19}$$

and then sketch a phase diagram for the solutions of this equation.

SOLUTION: Fist find the eigenvalues of matrix A above,

$$0 = \begin{vmatrix} (2-\lambda) & 3 \\ -3 & (2-\lambda) \end{vmatrix} = (\lambda-2)^2 + 9 \quad \Rightarrow \quad \lambda_\pm = 2 \pm 3i.$$

We then find the respective eigenvectors. The one corresponding to λ_+ is the solution of the homogeneous linear system with coefficients given by

$$\begin{bmatrix} 2-(2+3i) & 3 \\ -3 & 2-(2+3i) \end{bmatrix} = \begin{bmatrix} -3i & 3 \\ -3 & -3i \end{bmatrix} \to \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \to \begin{bmatrix} 1 & i \\ -1 & -i \end{bmatrix} \to \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}.$$

Therefore the eigenvector $\mathsf{v}_+ = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is given by

$$v_1 = -iv_2 \quad \Rightarrow \quad v_2 = 1, \quad v_1 = -i, \quad \Rightarrow \quad \mathsf{v}_+ = \begin{bmatrix} -i \\ 1 \end{bmatrix}, \quad \lambda_+ = 2 + 3i.$$

The second eigenvectors is the complex conjugate of the eigenvector found above, that is,

$$\mathsf{v}_- = \begin{bmatrix} i \\ 1 \end{bmatrix}, \quad \lambda_- = 2 - 3i.$$

Notice that

$$\mathsf{v}_\pm = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \pm \begin{bmatrix} -1 \\ 0 \end{bmatrix} i.$$

Hence, the real and imaginary parts of the eigenvalues and of the eigenvectors are given by

$$\alpha = 2, \qquad \beta = 3, \qquad \mathsf{a} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

So a real-valued expression for a fundamental set of solutions is given by

$$\mathsf{x}_1 = \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\cos(3t) - \begin{bmatrix} -1 \\ 0 \end{bmatrix}\sin(3t)\right) e^{2t} \quad \Rightarrow \quad \mathsf{x}_1 = \begin{bmatrix} \sin(3t) \\ \cos(3t) \end{bmatrix} e^{2t},$$
$$\mathsf{x}_2 = \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\sin(3t) + \begin{bmatrix} -1 \\ 0 \end{bmatrix}\cos(3t)\right) e^{2t} \quad \Rightarrow \quad \mathsf{x}_2 = \begin{bmatrix} -\cos(3t) \\ \sin(3t) \end{bmatrix} e^{2t}.$$

The phase diagram of these two fundamental solutions is given in Fig. 61 below. There is also a circle given in that diagram, corresponding to the trajectory of the vectors

$$\tilde{\mathsf{x}}_1 = \begin{bmatrix} \sin(3t) \\ \cos(3t) \end{bmatrix} \qquad \tilde{\mathsf{x}}_2 = \begin{bmatrix} -\cos(3t) \\ \sin(3t) \end{bmatrix}.$$

The trajectory of these vectors is a circle since their length is constant equal to one.

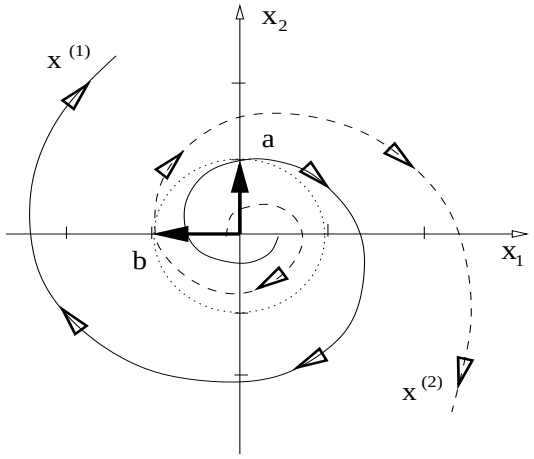

FIGURE 61. The graph of the fundamental solutions x_1 and x_2 of the Eq. (9.19).

In the particular case that the matrix A in Eq. (9.10) is 2×2, then any solutions is a linear combination of the solutions given in Eq. (9.18). That is, the general solution of given by

$$\mathsf{x}(t) = c_1\, \mathsf{x}_1(t) + c_2\, \mathsf{x}_2(t),$$

where

$$\mathsf{x}_1 = \big[\mathsf{a}\, \cos(\beta t) - \mathsf{b}\, \sin(\beta t)\big]\, e^{\alpha t}, \qquad \mathsf{x}_2 = \big[\mathsf{a}\, \sin(\beta t) + \mathsf{b}\, \cos(\beta t)\big]\, e^{\alpha t}.$$

We now do a qualitative study of the phase diagrams of the solutions for this case. We first fix the vectors a and b, and the plot phase diagrams for solutions having $\alpha > 0$, $\alpha = 0$, and $\alpha < 0$. These diagrams are given in Fig. 62. One can see that for $\alpha > 0$ the solutions spiral outward as t increases, and for $\alpha < 0$ the solutions spiral inwards to the origin as t increases..

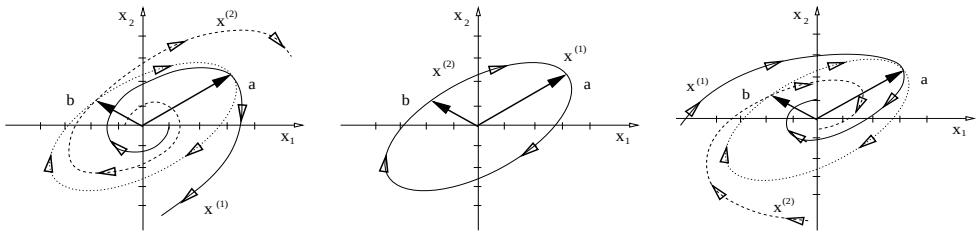

FIGURE 62. The graph of the fundamental solutions x_1 and x_2 (dashed line) of the Eq. (9.18) in the case of $\alpha > 0$, $\alpha = 0$, and $\alpha < 0$, respectively.

Finally, let us study the following cases: We fix $\alpha > 0$ and the vector a, and we plot the phase diagrams for solutions with two choices of the vector b, as shown in Fig. 63. It can then bee seen that the relative directions of the vectors a and b determines the rotation direction of the solutions as t increases.

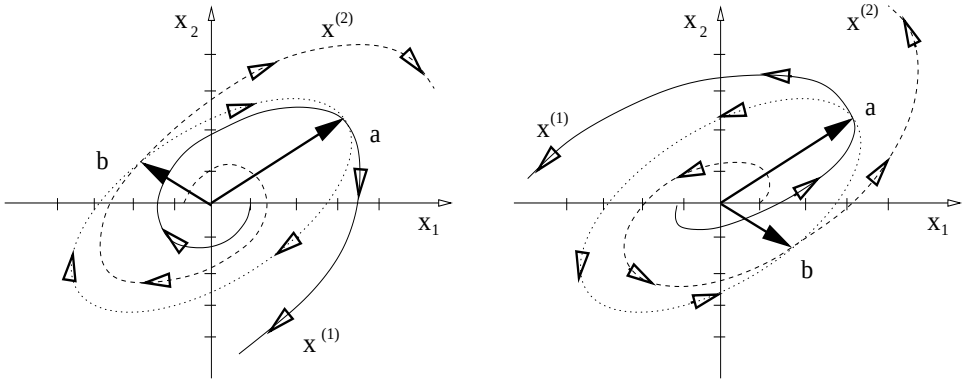

FIGURE 63. The graph of the fundamental solutions x_1 and x_2 (dashed line) of the Eq. (9.18) in the case of $\alpha > 0$, for a given vector a and for two choices of the vector b. The relative positions of the vectors a and b determines the rotation direction.

9.3.3. **Exercises.**

9.3.1.- Given the matrix $\mathsf{A} \in \mathbb{R}^{2,2}$ and vector $\mathsf{x}_0 \in \mathbb{R}^2$ below, find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solution of the initial value problem

$$\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}, \ \mathsf{x}_0 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

9.3.2.- Given the matrix A and the vector x_0 in Exercise **9.3.1** above, compute the operator valued function $e^{\mathsf{A}\,t}$ and verify that the solution x of the initial value problem given in Exercise **9.3.1** can be written as

$$\mathsf{x}(t) = e^{\mathsf{A}\,t}\,\mathsf{x}_0.$$

9.3.3.- Given the matrix $\mathsf{A} \in \mathbb{R}^{2,2}$ below, find all functions $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solutions of the differential equation

$$\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t),$$

where

$$\mathsf{A} = \begin{bmatrix} 1 & -1 \\ 5 & -3 \end{bmatrix}.$$

Since this matrix has complex eigenvalues, express the solutions as linear combination of real vector valued functions.

9.3.4.- Given the matrix $\mathsf{A} \in \mathbb{R}^{3,3}$ and vector $\mathsf{x}_0 \in \mathbb{R}^3$ below, find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^3$ solution of the initial value problem

$$\dot{\mathsf{x}}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where

$$\mathsf{A} = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathsf{x}_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

9.4. Normal operators

In this Section we introduce a particular type of linear operators called normal operators, which are defined on vector spaces having an inner product. Particular cases are Hermitian operators and unitary operators, which are widely used in physics. Rotations in space are examples of unitary operators on real vector spaces, while physical observables in quantum mechanics are examples of Hermitian operators. In this Section we restrict our description to finite dimensional inner product spaces. These definitions generalize the notions of unitary and Hermitian matrices already introduced in Chapter 2. We first describe the Riesz Representation Theorem, needed to verify that the notion of adjoint of a linear operator is well-defined. After reviewing the commutator of two operators we then introduce normal operators and discuss the particular cases of unitary and Hermitian operators. Finally we comment on the relations between these notions and the unitary and Hermitian matrices already introduced in Chapter 2.

The Riesz Representation Theorem is a statement concerning linear functionals on an inner product space. Given a vector space V over the scalar field $\mathbb{F}$, a **linear functional** is a scalar-valued linear function $f : V \to \mathbb{F}$, that is, for all $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and all a, $b \in \mathbb{F}$ holds $f(a\boldsymbol{x} + b\boldsymbol{y}) = a\, f(\boldsymbol{x}) + b\, f(\boldsymbol{y}) \in \mathbb{F}$. An example of a linear functional on $\mathbb{R}^3$ is the function

$$\mathbb{R}^3 \ni \mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto f(\mathsf{x}) = 3x_1 + 2x_2 + x_3 \in \mathbb{R}.$$

This function can be expressed in terms of the dot product in $\mathbb{R}^3$ as follows

$$f(\mathsf{x}) = \mathsf{u} \cdot \mathsf{x}, \qquad \mathsf{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

The Riesz Representation Theorem says that what we did in this example can be done in the general case. In an inner product space $\big(V, \langle\,,\rangle\big)$ every linear functional f can be expressed in terms of the inner product.

Theorem 9.4.1. *Consider a finite dimensional inner product space $\big(V, \langle\,,\rangle\big)$ over the scalar field $\mathbb{F}$. For every linear functional $f : V \to \mathbb{F}$ there exists a unique vector $\boldsymbol{u}_f \in V$ such that holds*

$$f(\boldsymbol{v}) = \langle \boldsymbol{u}_f, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in V.$$

Proof of Theorem 9.4.1: Introduce the set

$$N = \{\, \boldsymbol{v} \in V \;:\; f(\boldsymbol{v}) = 0 \,\} \subset V.$$

This set is the analogous to linear functionals of the null space of linear operators. Since f is a linear function the set N is a subspace of V. (Proof: Given two elements $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in N$ and two scalars a, $b \in \mathbb{F}$, holds that $f(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) = a\, f(\boldsymbol{v}_1) + b\, f(\boldsymbol{v}_2) = 0 + 0$, so $(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) \in N$.) Introduce the orthogonal complement of N, that is,

$$N^\perp = \{\boldsymbol{w} \in V \;:\; \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 0 \;\forall \boldsymbol{v} \in V\},$$

which is also a subspace of V. If $N^\perp = \{\boldsymbol{0}\}$, then $N = \big(N^\perp\big)^\perp = \big(\{\boldsymbol{0}\}\big)^\perp = V$. Since the null space of f is the whole vector space, the functional f is identically zero, so only for the choice $\boldsymbol{u}_f = \boldsymbol{0}$ holds $f(\boldsymbol{v}) = \langle \boldsymbol{0}, \boldsymbol{v} \rangle$ for all $\boldsymbol{v} \in V$.

In the case that $N^\perp \neq \{\boldsymbol{0}\}$ we now show that this space cannot be very big, in fact it has dimension one, as the following argument shows. Choose $\tilde{\boldsymbol{u}} \in N^\perp$ such that $f(\tilde{\boldsymbol{u}}) = 1$.

Then notice that for every $\boldsymbol{w} \in N^\perp$ the vector $\boldsymbol{w} - f(\boldsymbol{w})\tilde{\boldsymbol{u}}$ is trivially in $N^\perp$ but it is also in N, since

$$f\big(\boldsymbol{w} - f(\boldsymbol{w})\,\tilde{\boldsymbol{u}}\big) = f(\boldsymbol{w}) - f(\boldsymbol{w})\,f(\tilde{\boldsymbol{u}}) = f(\boldsymbol{w}) - f(\boldsymbol{w}) = 0.$$

A vector both in N and $N^\perp$ must vanish, so $\boldsymbol{w} = f(\boldsymbol{w})\,\tilde{\boldsymbol{u}}$. Then every vector in $N^\perp$ is proportional to $\tilde{\boldsymbol{u}}$, so $\dim N^\perp = 1$. This information is used to split any vector $\boldsymbol{v} \in V$ as follows $\boldsymbol{v} = a\,\tilde{\boldsymbol{u}} + \boldsymbol{x}$ where $\boldsymbol{x} \in V$ and $a \in \mathbb{F}$. It is clear that

$$f(\boldsymbol{v}) = f(a\,\tilde{\boldsymbol{u}} + \boldsymbol{x}) = a\,f(\tilde{\boldsymbol{u}}) + f(\boldsymbol{x}) = a\,f(\tilde{\boldsymbol{u}}) = a.$$

However, the function with values $g(\boldsymbol{v}) = \left\langle \dfrac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, \boldsymbol{v} \right\rangle$ has precisely the same values as f, since for all $\boldsymbol{v} \in V$ holds

$$g(\boldsymbol{v}) = \left\langle \frac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, \boldsymbol{v} \right\rangle = \left\langle \frac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, (a\,\tilde{\boldsymbol{u}} + \boldsymbol{x}) \right\rangle = \frac{a}{\|\tilde{\boldsymbol{u}}\|^2}\,\langle \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}} \rangle + \frac{1}{\|\tilde{\boldsymbol{u}}\|^2}\,\langle \tilde{\boldsymbol{u}}, \boldsymbol{x} \rangle = a.$$

Therefore, choosing $\boldsymbol{u}_f = \tilde{\boldsymbol{u}}/\|\tilde{\boldsymbol{u}}\|^2$, holds that

$$f(\boldsymbol{v}) = \langle \boldsymbol{u}_f, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in V.$$

Since $\dim N^\perp = 1$, the choice of $\boldsymbol{u}_f$ is unique. This establishes the Theorem. □

Given a linear operator defined on an inner product space, a new linear operator can be defined through an equation involving the inner product.

Proposition 9.4.2. *Let $\boldsymbol{T} \in L(V)$ be a linear operator on a finite-dimensional inner product space $\big(V, \langle\,,\,\rangle\big)$. There exists one and only one linear operator $\boldsymbol{T}^* \in L(V)$ such that*

$$\langle \boldsymbol{v},\, \boldsymbol{T}^*(\boldsymbol{u}) \rangle = \langle\, \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u} \rangle$$

holds for all vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$.

Given any linear operator $\boldsymbol{T}$ on a finite-dimensional inner product space, the operator $\boldsymbol{T}^*$ whose existence is guaranteed in Proposition 9.4.2 is called the **adjoint** of $\boldsymbol{T}$.

Proof of Proposition 9.4.2: We first establish the following statement: For every vector $\boldsymbol{u} \in V$ there exists a unique vector $\boldsymbol{w} \in V$ such that

$$\langle\, \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u} \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle \qquad \forall \boldsymbol{v} \in V. \tag{9.20}$$

The proof starts noticing that for a fixed $\boldsymbol{u} \in V$ the scalar-valued function $f_u : V \to \mathbb{F}$ given by $f_u(\boldsymbol{v}) = \langle u, \boldsymbol{T}(\boldsymbol{v}) \rangle$ is a linear functional. Therefore, the Riesz Representation Theorem 9.4.1 implies that there exists a unique vector $\boldsymbol{w} \in V$ such that $f_u(\boldsymbol{v}) = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$. This establishes that for every vector $\boldsymbol{u} \in V$ there exists a unique vector $\boldsymbol{w} \in V$ such that Eq. (9.20) holds. Now that this statement is proven we can define a map, that we choose to denote as $\boldsymbol{T}^* : V \to V$, given by $\boldsymbol{u} \mapsto \boldsymbol{T}^*(\boldsymbol{u}) = \boldsymbol{w}$. We now show that this map $\boldsymbol{T}^*$ is linear. Indeed, for all $\boldsymbol{u}_1$, $\boldsymbol{u}_2 \in V$ and all a, $b \in \mathbb{F}$ holds

$$\begin{aligned}
\langle \boldsymbol{v},\, \boldsymbol{T}^*(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) \rangle &= \langle\, \boldsymbol{T}(\boldsymbol{v}), (a\boldsymbol{u}_1 + b\boldsymbol{u}_2) \rangle \qquad \forall \boldsymbol{v} \in V, \\
&= a\,\langle\, \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}_1 \rangle + b\,\langle\, \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}_2) \rangle \\
&= a \langle \boldsymbol{v},\, \boldsymbol{T}^*(\boldsymbol{u}_1) \rangle + b \langle \boldsymbol{v},\, \boldsymbol{T}^*(\boldsymbol{u}_2) \rangle \\
&= \big\langle \boldsymbol{v},\, \big[a\,\boldsymbol{T}^*(\boldsymbol{u}_1) + b\,\boldsymbol{T}^*(\boldsymbol{u}_2)\big] \big\rangle \qquad \forall \boldsymbol{v} \in V,
\end{aligned}$$

hence $\boldsymbol{T}^*(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) = a\,\boldsymbol{T}^*(\boldsymbol{u}_1) + b\,\boldsymbol{T}^*(\boldsymbol{u}_2)$. This establishes the Proposition. □

The next result relates the adjoint of a linear operator with the concept of the adjoint of a square matrix introduced in Sect. 2.2. Recall that given a basis in the vector space, every linear operator has associated a unique square matrix. Let us use the notation $[\boldsymbol{T}]$ and $[\boldsymbol{T}^*]$ for the matrices on a given basis of the operators $\boldsymbol{T}$ and $\boldsymbol{T}^*$, respectively.

Proposition 9.4.3. *Let $\big(V, \langle\,,\rangle\big)$ be a finite-dimensional vector space, let $\mathcal{V}$ be an orthonormal basis of V, and let $[\boldsymbol{T}]$ be the matrix of the linear operator $\boldsymbol{T} \in L(V)$ in the basis $\mathcal{V}$. Then, the matrix of the adjoint operator $\boldsymbol{T}^*$ in the basis $\mathcal{V}$ is given by $[\boldsymbol{T}^*] = [\boldsymbol{T}]^*$.*

Proposition 9.4.3 says that the matrix of the adjoint operator is the adjoint of the matrix of the operator, however this is true only in the case that the basis used to compute the respective matrices is orthonormal. If the basis is not orthonormal, the relation between the matrices $[\boldsymbol{T}]$ and $[\boldsymbol{T}^*]$ is more involved.

Proof of Proposition 9.4.3: Let $\mathcal{V} = \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ be an orthonormal basis of V, that is,

$$\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j. \end{cases}$$

The components of two arbitrary vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$ in the basis $\mathcal{V}$ is denoted as follows

$$\boldsymbol{u} = \sum_i u_i \boldsymbol{e}_i, \qquad \boldsymbol{v} = \sum_i v_i \boldsymbol{e}_i.$$

The action of the operator $\boldsymbol{T}$ can also be decomposed in the basis $\mathcal{V}$ as follows

$$\boldsymbol{T}(\boldsymbol{e}_j) = \sum_i [\boldsymbol{T}]_{ij} \boldsymbol{e}_i, \qquad [\boldsymbol{T}]_{ij} = [\boldsymbol{T}(\boldsymbol{e}_j)]_i.$$

We use the same notation for the adjoint operator, that is,

$$\boldsymbol{T}^*(\boldsymbol{e}_j) = \sum_i [\boldsymbol{T}^*]_{ij} \boldsymbol{e}_i, \qquad [\boldsymbol{T}^*]_{ij} = [\boldsymbol{T}^*(\boldsymbol{e}_j)]_i.$$

The adjoint operator is defined such that the equation $\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u}) \rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u} \rangle$ holds for all $\boldsymbol{u}$, $\boldsymbol{v} \in V$. This equation can be expressed in terms of components in the basis $\mathcal{V}$ as follows

$$\sum_{ijk} \big\langle v_i \boldsymbol{e}_i, u_j [\boldsymbol{T}^*(\boldsymbol{e}_j)]_k \boldsymbol{e}_k \big\rangle = \sum_{ijk} \big\langle v_i [\boldsymbol{T}(\boldsymbol{e}_i)]_k \boldsymbol{e}_k, u_j \boldsymbol{e}_j \big\rangle,$$

that is,

$$\sum_{ijk} \overline{v}_i u_j [\boldsymbol{T}^*]_{kj} \langle \boldsymbol{e}_i, \boldsymbol{e}_k \rangle = \sum_{ijk} \overline{v}_i \overline{[\boldsymbol{T}]}_{ki} u_j \langle \boldsymbol{e}_k, \boldsymbol{e}_j \rangle.$$

Since the basis $\mathcal{V}$ is orthonormal we obtain the equation

$$\sum_{ij} \overline{v}_i u_j [\boldsymbol{T}^*]_{ij} = \sum_{ijk} \overline{v}_i \overline{[\boldsymbol{T}]}_{ji} u_j,$$

which holds for all vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$, so we conclude

$$[\boldsymbol{T}^*]_{ij} = \overline{[\boldsymbol{T}]}_{ji} \quad \Leftrightarrow \quad [\boldsymbol{T}^*] = \overline{[\boldsymbol{T}]}^T \quad \Leftrightarrow \quad [\boldsymbol{T}^*] = [\boldsymbol{T}]^*.$$

This establishes the Proposition. □

EXAMPLE 9.4.1: Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$. Find the adjoint of the linear operator $\boldsymbol{T}$ with matrix in the standard basis of $\mathbb{C}^3$ given by

$$[\boldsymbol{T}(\boldsymbol{x})] = \begin{bmatrix} x_1 + 2ix_2 + ix_3 \\ ix_1 - x_3 \\ x_1 - x_2 + ix_3 \end{bmatrix}, \qquad [\boldsymbol{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

SOLUTION: The matrix of this operator in the standard basis of $\mathbb{C}^3$ is given by

$$[\boldsymbol{T}] = \begin{bmatrix} 1 & 2i & i \\ i & 0 & -1 \\ 1 & -1 & i \end{bmatrix}.$$

Since the standard basis is an orthonormal basis with respect to the dot product, Proposition 9.4.3 implies that

$$[\boldsymbol{T}^*] = [\boldsymbol{T}]^* = \begin{bmatrix} 1 & 2i & i \\ i & 0 & -1 \\ 1 & -1 & i \end{bmatrix}^* = \begin{bmatrix} 1 & -i & 1 \\ -2i & 0 & -1 \\ -i & -1 & -i \end{bmatrix} \quad \Rightarrow \quad [\boldsymbol{T}^*(\boldsymbol{x})] = \begin{bmatrix} x_1 - ix_2 + x_3 \\ -2ix_1 - x_3 \\ -ix_1 - x_2 - ix_3 \end{bmatrix}.$$

◁

Recall now that the commutator of two linear operators $\boldsymbol{T}, \boldsymbol{S} \in L(V)$ is the linear operator $[\boldsymbol{T}, \boldsymbol{S}] \in L(V)$ given by

$$[\boldsymbol{T}, \boldsymbol{S}](\boldsymbol{u}) = \boldsymbol{T}(\boldsymbol{S}(\boldsymbol{u})) - \boldsymbol{S}(\boldsymbol{T}(\boldsymbol{u})) \qquad \forall \boldsymbol{u} \in V.$$

Two operators $\boldsymbol{T}, \boldsymbol{S} \in L(V)$ are said to commute iff their commutator vanishes, that is, $[\boldsymbol{T}, \boldsymbol{S}] = \boldsymbol{0}$. Examples of operators that commute are two rotations on the plane. Examples of operators that do not commute are two arbitrary rotations in space.

Definition 9.4.4. *A linear operator $\boldsymbol{T}$ defined on a finite-dimensional inner product space $\big(V, \langle\,,\rangle\big)$ is called a* **normal operator** *iff holds $[\boldsymbol{T}, \boldsymbol{T}^*] = \boldsymbol{0}$, that is, the operator commutes with its adjoint.*

An interesting characterization of normal operators is the following: A linear operator $\boldsymbol{T}$ on an inner product space is normal iff $\|\boldsymbol{T}(\boldsymbol{u})\| = \|\boldsymbol{T}^*(\boldsymbol{u})\|$ holds for all $\boldsymbol{u} \in V$. Normal operators are particularly important because for these operators hold the Spectral Theorem, which we study in Chapter 9.

Two particular cases of normal operators are often used in physics. A linear operator $\boldsymbol{T}$ on an inner product space is called a **unitary operator** iff $\boldsymbol{T}^* = \boldsymbol{T}^{-1}$, that is, the adjoint is the inverse operator. Unitary operators are normal operators, since

$$\boldsymbol{T}^* = \boldsymbol{T}^{-1} \quad \Rightarrow \quad \begin{cases} \boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{I}, \\ \boldsymbol{T}^*\boldsymbol{T} = \boldsymbol{I}, \end{cases} \quad \Rightarrow \quad [\boldsymbol{T}, \boldsymbol{T}^*] = \boldsymbol{0}.$$

Unitary operators preserve the length of a vector, since

$$\|\boldsymbol{v}\|^2 = \langle \boldsymbol{v}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{T}^{-1}(\boldsymbol{T}(\boldsymbol{v})) \rangle = \langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{T}(\boldsymbol{v})) \rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{T}(\boldsymbol{v}) \rangle = \|\boldsymbol{T}(\boldsymbol{v})\|^2.$$

Unitary operators defined on a complex inner product space are particularly important in quantum mechanics. The particular case of unitary operators on a real inner product space are called **orthogonal operators**. So, orthogonal operators do not change the length of a vector. Examples of orthogonal operators are rotations in $\mathbb{R}^3$.

A linear operator $\boldsymbol{T}$ on an inner product space is called an **Hermitian operator** iff $\boldsymbol{T}^* = \boldsymbol{T}$, that is, the adjoint is the original operator. This definition agrees with the definition of Hermitian matrices given in Chapter 2.

EXAMPLE 9.4.2: Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$ and the linear operator $\boldsymbol{T}$ with matrix in the standard basis of $\mathbb{C}^3$ given by

$$[\boldsymbol{T}(\boldsymbol{x})] = \begin{bmatrix} x_1 - ix_2 + x_3 \\ ix_1 - x_3 \\ x_1 - x_2 + x_3 \end{bmatrix}, \qquad [\boldsymbol{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Show that $\boldsymbol{T}$ is Hermitian.

SOLUTION: We need to compute the adjoint of $\boldsymbol{T}$. The matrix of this operator in the standard basis of $\mathbb{C}^3$ is given by

$$[\boldsymbol{T}] = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Since the standard basis is an orthonormal basis with respect to the dot product, Proposition 9.4.3 implies that

$$[\boldsymbol{T}^*] = [\boldsymbol{T}]^* = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}^* = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} = [\boldsymbol{T}].$$

Therefore, $\boldsymbol{T}^* = \boldsymbol{T}$. ◁

9.4.1. **Exercises.**

9.4.1.- .

9.4.2.- .

Chapter 10. Appendix

10.1. Review exercises

Chapter 1: Linear systems

1.- Consider the linear system

$$2x_1 + 3x_2 - x_3 = 6$$
$$-x_1 - x_2 + 2x_3 = -2$$
$$x_1 + 2x_3 = 2$$

(a) Use Gauss operations to find the reduced echelon form of the augmented matrix for this system.

(b) Is this system consistent? If "yes," find all the solutions.

2.- Find all the solutions x to the linear system $\mathsf{Ax} = \mathsf{b}$ and express them in vector form, where

$$\mathsf{A} = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 1 & 8 \\ 1 & -1 & 1 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

3.- Consider the matrix and the vector

$$\mathsf{A} = \begin{bmatrix} 1 & -2 & 7 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.$$

Is the vector b a linear combination of the column vectors of A?

4.- Let s be a real number, and consider the system

$$sx_1 - 2sx_2 = -1,$$
$$3x_1 + 6sx_2 = 3.$$

(a) Determine the values of the parameter s for which the system above has a unique solution.

(b) For all the values of s such that the system above has a unique solution, find that solution.

5.- Find the values of k such that the system below has no solution; has one solution; has infinitely many solutions;

$$kx_1 + x_2 = 1$$
$$x_1 + kx_2 = 1.$$

6.- Find a condition on the components of vector b such that the system $\mathsf{Ax} = \mathsf{b}$ is consistent, where

$$\mathsf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -6 \\ 3 & 1 & -7 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

7.- Find the general solution to the homogeneous linear system with coefficient matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 3 & -1 & 5 \\ 2 & 1 & 3 & 0 \\ 3 & 2 & 4 & 1 \end{bmatrix},$$

and write this general solution in vector form.

8.- (a) Find a value of the constants h and k such that the non-homogeneous linear system below is consistent and has one free variable.

$$x_1 + h\,x_2 + 5x_3 = 1,$$
$$x_2 - 2x_3 = k,$$
$$x_1 + 3x_2 - 3x_3 = 5.$$

(b) Using the value of the constants h and k found in part (a), find the general solution to the system given in part (a).

9.- (a) Find the general solution to the system below and write it in vector form,

$$x_1 + 2x_2 - x_3 = 2,$$
$$3x_1 + 7x_2 - 3x_3 = 7,$$
$$x_1 + 4x_2 - x_3 = 4.$$

(b) Sketch a graph on $\mathbb{R}^3$ of the general solution found in part (a).

Chapter 2: Matrix algebra

1.- Consider the vectors

$$\mathsf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and the linear function $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$T(\mathsf{u}) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad T(\mathsf{v}) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Find the matrix $\mathsf{A} = [T(\mathsf{e}_1), T(\mathsf{e}_2)]$ of the linear transformation, where

$$\mathsf{e}_1 = \frac{1}{2}(\mathsf{u} + \mathsf{v}), \quad \mathsf{e}_2 = \frac{1}{2}(\mathsf{u} - \mathsf{v}).$$

Show your work.

2.- Find the matrix for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ representing a reflection on the plane along the vertical axis followed by a rotation by $\theta = \pi/3$ counterclockwise.

3.- Which of the following matrices below is equal to $(\mathsf{A} + \mathsf{B})^2$ for every square matrices A and B?

$$(\mathsf{B} + \mathsf{A})^2,$$
$$\mathsf{A}^2 + 2\mathsf{AB} + \mathsf{B}^2,$$
$$(\mathsf{A} + \mathsf{B})(\mathsf{B} + \mathsf{A}),$$
$$\mathsf{A}^2 + \mathsf{AB} + \mathsf{BA} + \mathsf{B}^2,$$
$$\mathsf{A}(\mathsf{A} + \mathsf{B}) + (\mathsf{A} + \mathsf{B})\mathsf{B}.$$

4.- Find a matrix A solution of the matrix equation

$$\mathsf{AB} + 2\,\mathsf{I}_2 = \begin{bmatrix} 5 & 4 \\ -2 & 3 \end{bmatrix},$$

where

$$\mathsf{B} = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}.$$

5.- Consider the matrix

$$\mathsf{A} = \begin{bmatrix} 5 & 3 & s \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix}.$$

Find the value(s) of the constant s such that the matrix A is invertible.

6.- Let A be an $n \times n$ matrix, D be and $m \times m$ matrix, and C be and $m \times n$ matrix. Assume that both A and D are invertible matrices and denote by A^{-1}, D^{-1} their respective inverse matrices. Let M be the $(n + m) \times (n + m)$ matrix

$$\mathsf{M} = \begin{bmatrix} \mathsf{A} & 0 \\ \mathsf{C} & \mathsf{D} \end{bmatrix}.$$

Find an $m \times n$ matrix X (in terms of any of the matrices A, D, A^{-1}, D^{-1}, and C) such that M is invertible and the inverse is given by

$$\mathsf{M}^{-1} = \begin{bmatrix} \mathsf{A}^{-1} & 0 \\ \mathsf{X} & \mathsf{D}^{-1} \end{bmatrix}.$$

7.- Consider the matrix and the vector

$$\mathsf{A} = \begin{bmatrix} 1 & -2 & 7 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.$$

(a) Does vector b belong to the $R(\mathsf{A})$?
(b) Does vector b belong to the $N(\mathsf{A})$?

8.- Consider the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 6 & -7 \\ -2 & 3 & 2 & 0 \\ 0 & -1 & -2 & 2 \end{bmatrix}.$$

Find $N(\mathsf{A})$ and the $R(\mathsf{A}^T)$.

9.- Consider the matrix

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 7 \\ 6 & 9 & 12 \end{bmatrix}.$$

(a) Find the LU-factorization of A.
(b) Use the LU-factorization above to find the solutions x_1 and x_2 of the systems $\mathsf{Ax} = \mathsf{e}_1$ and $\mathsf{Ax}_2 = \mathsf{e}_2$, where

$$\mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

(c) Use the LU-factorization above to find A^{-1}.

Chapter 3: Determinants

1.- Find the determinant of the matrices

$$\mathsf{A} = \begin{bmatrix} 1+i & -3i \\ -4i & 1-2i \end{bmatrix},$$

$$\mathsf{B} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & -2 & 3 & -2 \\ 0 & 0 & -1 & -3 \\ -2 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathsf{C} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

2.- Given matrix A below, find the cofactors matrix C, and explicitly show that $\mathsf{A}\,\mathsf{C}^T = \det(\mathsf{A})\,\mathsf{I}_3$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 4 & 0 & 1 \\ 2 & 1 & 3 \end{bmatrix}.$$

3.- Given matrix A below, find the coefficients $(\mathsf{A}^{-1})_{13}$ and $(\mathsf{A}^{-1})_{23}$ of the inverse matrix A^{-1}, where

$$\mathsf{A} = \begin{bmatrix} 5 & 3 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix}.$$

4.- Find the change in the area of the parallelogram formed by the vectors

$$\mathsf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

when this parallelogram is transformed under the following linear transformation, $\mathsf{A} : \mathbb{R}^2 \to \mathbb{R}^2$,

$$\mathsf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

5.- Find the volume of the parallelepiped formed by the vectors

$$\mathsf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \quad \mathsf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

6.- Consider the matrix

$$\mathsf{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}.$$

Find the values of the scalar λ such that the matrix $(\mathsf{A} - \lambda\,\mathsf{I}_2)$ is not invertible.

7.- Prove the following: If there exists an integer $k \geqslant 1$ such that $\mathsf{A} \in \mathbb{F}^{n,n}$ satisfies $\mathsf{A}^k = 0$, then $\det(\mathsf{A}) = 0$.

8.- Assume that matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ satisfies the equation $\mathsf{A}^2 = \mathsf{I}_n$. Find all possible values of $\det(\mathsf{A})$.

9.- Use Cramer's rule to find the solution of the linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 4 & -1 \\ 1 & 1 & 1 \\ 2 & 0 & 3 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Chapter 4: Vector spaces

1.- Determine which of the following subsets of $\mathbb{R}^{3,3}$ are subspaces:

(a) The symmetric matrices.
(b) The skew-symmetric matrices.
(c) The matrices A with $\mathsf{A}^2 = \mathsf{A}$.
(d) The matrices A with $\mathrm{tr}\,(\mathsf{A}) = 0$.
(e) The matrices A with $\det(\mathsf{A}) = 0$.

In the case that the set is a subspace, find a basis of this subspace.

2.- Find the dimension and give a basis of the subspace $W \subset \mathbb{R}^3$ given by

$$\left\{ \begin{bmatrix} -a+b+c-3d \\ b+3c-d \\ a+2b+8c \end{bmatrix} \text{ with } a,b,c,d \in \mathbb{R} \right\}.$$

3.- Find the dimension of both the null space and the range space of the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 1 & 1 & 5 & 2 \\ 2 & 1 & 4 & 7 & 3 \\ 0 & -1 & 2 & -3 & -1 \end{bmatrix}.$$

4.- Consider the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & -1 & 5 \\ 0 & 1 & -2 \\ 1 & 3 & -3 \end{bmatrix}.$$

(a) Find a basis for the null space of A.
(b) Find a basis for the subspace in $\mathbb{R}^3$ consisting of all vectors $\mathsf{b} \in \mathbb{R}^3$ such that the linear system $\mathsf{Ax} = \mathsf{b}$ is consistent.

5.- Show whether the following statement is true or false: Given a vector space V, if the set

$$\{\mathsf{v}_1, \mathsf{v}_2, \mathsf{v}_3\} \subset V$$

is linearly independent, then so is the set

$$\{\mathsf{w}_1,\ \mathsf{w}_2,\ \mathsf{w}_3\},$$

where

$$\mathsf{w}_1 = (\mathsf{v}_1 + \mathsf{v}_2),$$
$$\mathsf{w}_2 = (\mathsf{v}_1 + \mathsf{v}_3),$$
$$\mathsf{w}_3 = (\mathsf{v}_2 + \mathsf{v}_3).$$

6.- Show that the set $U \subset \mathbb{P}_3$ given by all polynomials satisfying the condition

$$\int_0^1 \boldsymbol{p}(x)\, dx = 0$$

is a subspace of $\mathbb{P}_3$. Find a basis for U.

7.- Determine whether the set $U \subset \mathbb{P}_2$ of all polynomials of the form

$$\boldsymbol{p}(x) = a + ax + ax^2$$

with $a \in \mathbb{F}$, is a subspace of $\mathbb{P}_2$.

Chapter 5: Linear transformations

1.- Consider the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 6 & -7 \\ -2 & 3 & 2 & 0 \\ 0 & -1 & -2 & 2 \end{bmatrix}.$$

(1) Find a basis for $R(A)$.
(2) Find a basis for $N(A)$.
(3) Consider the linear transformation $\mathsf{A} : \mathbb{R}^4 \to \mathbb{R}^3$ determined by A. Is it injective? Is it surjective? Justify your answers.

2.- Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation given by

$$[T(\mathsf{x}_s)]_{\tilde{s}} = \begin{bmatrix} 2x_1 + 6x_2 - 2x_3 \\ 3x_1 + 8x_2 + 2x_3 \end{bmatrix}_s,$$

$$[\mathsf{x}]_s = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

where $\mathcal{S}$ and $\tilde{\mathcal{S}}$ are the standard bases in $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively.

(a) Is T injective? Is T surjective?
(b) Find all solutions of the linear system $T(\mathsf{x}_s) = \mathsf{b}_{\tilde{s}}$, where

$$\mathsf{b}_{\tilde{s}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

(c) Is the set of all solutions found in part (b) a subspace of $\mathbb{R}^3$?

3.- Let $\mathsf{A} : \mathbb{C}^3 \to \mathbb{C}^4$ be the linear transformation

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & i \\ 0 & i & 1 \\ 1-i & 0 & 1+i \\ i & 1 & 0 \end{bmatrix}.$$

Find a basis for $N(\mathsf{A})$ and $R(\mathsf{A})$.

4.- Let $\boldsymbol{D} : \mathbb{P}_3 \to \mathbb{P}_3$ be the differentiation operator,

$$\boldsymbol{D}(\boldsymbol{p})(x) = \frac{d\boldsymbol{p}}{dx}(x),$$

and $\boldsymbol{I} : \mathbb{P}_3 \to \mathbb{P}_3$ the identity operator. Let $\mathcal{S} = (1,\ x,\ x^2,\ x^3)$ be the standard ordered basis of $\mathbb{P}_3$. Show that the matrix of the operator

$$(\boldsymbol{I} - \boldsymbol{D}^2) : \mathbb{P}_3 \to \mathbb{P}_3$$

in the basis $\mathcal{S}$ is invertible.

Chapter 6: Inner product spaces

1.- Let $\big(V, \langle\,,\rangle\big)$ be a real inner product space. Show that $(\boldsymbol{x}-\boldsymbol{y}) \perp (\boldsymbol{x}+\boldsymbol{y})$ iff $\|\boldsymbol{x}\| = \|\boldsymbol{y}\|$.

2.- Consider the inner product space given by $\big(\mathbb{F}^n, \cdot\big)$. Prove that for every matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ holds

$$\mathsf{x} \cdot (\mathsf{A}\mathsf{y}) = (\mathsf{A}^*\mathsf{x}) \cdot \mathsf{y}.$$

3.- A matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ is called **unitary** iff

$$\mathsf{A}\,\mathsf{A}^* = \mathsf{A}^*\mathsf{A} = \mathsf{I}_n.$$

Show that a unitary matrix does not change the norm of a vector in the inner product space $\big(\mathbb{F}^n, \cdot\big)$, that is, for all $\mathsf{x} \in \mathbb{F}^n$ and all unitary matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ holds

$$\|\mathsf{A}\mathsf{x}\| = \|\mathsf{x}\|.$$

4.- Find all vectors in the inner product space $\big(\mathbb{R}^4, \cdot\big)$ perpendicular to both

$$\mathsf{v} = \begin{bmatrix} 1 \\ 4 \\ 4 \\ 1 \end{bmatrix} \text{ and } \mathsf{u} = \begin{bmatrix} 2 \\ 9 \\ 8 \\ 2 \end{bmatrix}.$$

5.- Consider the inner product space given by $\big(\mathbb{C}^3, \cdot\big)$ and the subspace $W \subset \mathbb{C}^3$ spanned by the vectors

$$\mathsf{u} = \begin{bmatrix} 1+i \\ 1 \\ i \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} -1 \\ 0 \\ 2-i \end{bmatrix}.$$

(a) Use the Gram-Schmidt method to find an orthonormal basis for W.

(b) Extend the orthonormal basis of W into an orthonormal basis for $\mathbb{C}^3$.

Chapter 7: Approximation methods

1.- .

2.- .

Chapter 8: Normed spaces

1.- .

2.- .

Chapter 9: Spectral decomposition

1.- .

2.- .

10.2. Practice Exams

Instructions to use the Practice Exams. The idea of these lecture notes is to help anyone interested in learning linear algebra. More often than not such person is a college student. An unfortunate aspect of our education system is that students must pass an exam to prove they understood the ideas of a given subject. To prepare the student for such exam is the purpose of the practice exams in this Section.

These practice exams once were actual exams taken by student in previous courses. These exams can be useful to other students if they are used properly; one way is the following: Study the course material first, do all the exercises at the end of every Section; do the review problems in Section 10.1; then and only then take the first practice exam and do it. Think of it as an actual exam. Do not look at the notes or any other literature. Do the whole exam. Watch your time. You have only two hours to do it. After you finish, you can grade yourself. You have the solutions to the exam at the end of the Chapter. Never, ever look at the solutions before you finish the exam. If you do it, the practice exam is worthless. Really, worthless; you will not have the solutions when you do the actual exam. The story does not finish here. Pay close attention at the exercises you did wrong, if any. Go back to the class material and do extra problems on those subjects. Review all subjects you think you are not well prepared. Then take the second practice exam. Follow a similar procedure. Review the subject related to the practice exam exercises you had difficulties to solve. Then, do the next practice exam. You have three of them. After doing the last one, you should have a good idea of what your actual grade in the class should be.

Practice Exam 1. (Two hours.)

1. Consider the matrix $\mathsf{A} = \begin{bmatrix} 5 & 3 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix}$. Find the coefficients $(\mathsf{A}^{-1})_{21}$ and $(\mathsf{A}^{-1})_{32}$ of the matrix A^{-1}, that is, of the inverse matrix of A. Show your work.

2. (a) Find $k \in \mathbb{R}$ such that the volume of the parallelepiped formed by the vectors below is equal to 4, where

$$\mathsf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \quad \mathsf{v}_3 = \begin{bmatrix} k \\ 1 \\ 1 \end{bmatrix}$$

(b) Set $k = 1$ and define the matrix $\mathsf{A} = [\mathsf{v}_1, \mathsf{v}_2, \mathsf{v}_3]$. Matrix A determines the linear transformation $\mathsf{A} : \mathbb{R}^3 \to \mathbb{R}^3$. Is this linear transformation injective (one-to-one)? Is it surjective (onto)?

3. Determine whether the subset $V \subset \mathbb{R}^3$ is a subspace, where

$$V = \left\{ \begin{bmatrix} -a + b \\ a - 2b \\ a - 7b \end{bmatrix} \quad \text{with} \quad a, b \in \mathbb{R} \right\}.$$

If the set is a subspace, find an orthogonal basis in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

4. True or False: (Justify your answers.)

(a) If the set of columns of $\mathsf{A} \in \mathbb{F}^{m,n}$ is a linearly independent set, then $\mathsf{Ax} = \mathsf{b}$ has exactly one solution for every $\mathsf{b} \in \mathbb{F}^m$.

(b) The set of column vectors of an 5×7 is never linearly independent.

5. Consider the vector space $\mathbb{R}^2$ with the standard basis $\mathcal{S}$ and let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation

$$[\boldsymbol{T}]_{ss} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}_{ss}.$$

Find $[\boldsymbol{T}]_{bb}$, the matrix of $\boldsymbol{T}$ in the basis $\mathcal{B}$, where $\mathcal{B} = \left\{ [\mathsf{b}_1]_s = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_s, [\mathsf{b}_2]_s = \begin{bmatrix} 1 \\ -2 \end{bmatrix}_s \right\}$.

6. Consider the linear transformations $\boldsymbol{T} : \mathbb{R}^3 \to \mathbb{R}^2$ and $\boldsymbol{S} : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\left[\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{s_3} \Big) \right]_{s_2} = \begin{bmatrix} x_1 - x_2 + x_3 \\ -x_1 + 2x_2 + x_3 \end{bmatrix}_{s_2}, \qquad \left[\boldsymbol{S}\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{s_3} \Big) \right]_{s_3} = \begin{bmatrix} 3x_3 \\ 2x_2 \\ x_1 \end{bmatrix}_{s_3},$$

where $\mathcal{S}_3$ and $\mathcal{S}_2$ are the standard basis of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively.

(a) Find a matrix $[\boldsymbol{T}]_{s_3 s_2}$ and the matrix $[\boldsymbol{S}]_{s_3 s_3}$. Show your work.

(b) Find the matrix of the composition $\boldsymbol{T} \circ \boldsymbol{S} : \mathbb{R}^3 \to \mathbb{R}^2$ in the standard basis, that is, find $[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2}$.

(c) Is $\boldsymbol{T} \circ \boldsymbol{S}$ injective (one-to-one)? Is $\boldsymbol{T} \circ \boldsymbol{S}$ surjective (onto)? Justify your answer.

7. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}$ and the vector $\mathsf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$.

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the matrix equation $\mathsf{A}\mathsf{x} = \mathsf{b}$.
(b) Verify whether the vector $\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}$ belong to the space $R(\mathsf{A})^{\perp}$? Justify your answers.

8. Consider the matrix $\mathsf{A} = \begin{bmatrix} -1/2 & -3 \\ 1/2 & 2 \end{bmatrix}$.
(a) Show that matrix A is diagonalizable.
(b) Using that A is diagonalizable, find the $\lim_{k\to\infty} \mathsf{A}^k$.

9. Let $\big(V, \langle\,,\rangle\big)$ be an inner product space with inner product norm $\|\ \|$. Let $\boldsymbol{T}: V \to V$ be a linear transformation and $\boldsymbol{x}$, $\boldsymbol{y} \in V$ be vectors satisfying the following conditions:

$$\boldsymbol{T}(\boldsymbol{x}) = 2\,\boldsymbol{x}, \quad \boldsymbol{T}(\boldsymbol{y}) = -3\,\boldsymbol{y}, \quad \|\boldsymbol{x}\| = 1/3, \quad \|\boldsymbol{y}\| = 1, \quad \boldsymbol{x} \perp \boldsymbol{y}.$$

(a) Compute $\|\boldsymbol{v}\|$ for the vector $\boldsymbol{v} = 3\boldsymbol{x} - \boldsymbol{y}$.
(b) Compute $\|\boldsymbol{T}(\boldsymbol{v})\|$ for the vector $\boldsymbol{v}$ given above.

10. Consider the matrix $\mathsf{A} = \begin{bmatrix} 2 & -1 & 2 \\ 0 & 1 & h \\ 0 & 0 & 2 \end{bmatrix}$.
(a) Find all eigenvalues of matrix A and their corresponding algebraic multiplicities.
(b) Find the value(s) of the real number h such that the matrix A above has a two-dimensional eigenspace, and find a basis for this eigenspace.

Practice Exam 2. (Two hours.)

1. Consider the matrix $\mathsf{A} = \begin{bmatrix} -2 & 3 & -1 \\ 1 & 2 & -1 \\ -2 & -1 & 1 \end{bmatrix}$. Find the coefficients $(\mathsf{A}^{-1})_{13}$ and $(\mathsf{A}^{-1})_{21}$ of the inverse matrix of A. Show your work.

2. Consider the vector space $\mathbb{P}_3([0,1])$ with the inner product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\,dx.$$

Given the set $\mathcal{U} = \{\boldsymbol{p}_1 = x^2,\ \boldsymbol{p}_2 = x^3\}$, find an orthogonal basis for the subspace $U = \mathrm{Span}(\mathcal{U})$ using the Gram-Schmidt method on the set $\mathcal{U}$ starting with the vector $\boldsymbol{p}_1$.

3. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 3 & 1 & 1 \\ 2 & 6 & 3 & 0 \\ 3 & 9 & 5 & -1 \end{bmatrix}$.

(a) Verify that the vector $\mathsf{v} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ -2 \end{bmatrix}$ belongs to the null space of A.

(b) Extend the set $\{\mathsf{v}\}$ into a basis of the null space of A.

4. Use Cramer's rule to find the solution to the linear system

$$\begin{aligned} 2x_1 + x_2 - x_3 &= 0 \\ x_1 + x_3 &= 1 \\ x_1 + 2x_2 + 3x_3 &= 0. \end{aligned}$$

5. Let $\mathcal{S}_3$ and $\mathcal{S}_2$ be standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively, and consider the linear transformation $\boldsymbol{T}\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$\left[\boldsymbol{T}\Big(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{s_3}\Big)\right]_{s_2} = \begin{bmatrix} -x_1 + 2x_2 - x_3 \\ x_1 + x_3 \end{bmatrix}_{s_2},$$

and introduce the bases $\mathcal{U} \subset \mathbb{R}^3$ and $\mathcal{V} \subset \mathbb{R}^2$ given by

$$\mathcal{U} = \Big\{ [\mathsf{u}_1]_{s_3} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_{s_3},\ [\mathsf{u}_2]_{s_3} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_{s_3},\ [\mathsf{u}_3]_{s_3} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_{s_3} \Big\},$$

$$\mathcal{V} = \Big\{ [\mathsf{v}_1]_{s_2} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{s_2},\ [\mathsf{v}_2]_{s_2} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}_{s_2} \Big\}.$$

Find the matrices $[\boldsymbol{T}]_{s_3 s_2}$ and $[\boldsymbol{T}]_{uv}$. Show your work.

6. Consider the inner product space $\big(\mathbb{R}^{2,2}, \langle\,,\rangle_F\big)$ and the subspace

$$W = \mathrm{Span}\Big\{ \mathsf{E}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ \mathsf{E}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \Big\} \subset \mathbb{R}^{2,2}.$$

Find a basis for $W^\perp$, the orthogonal complement of W.

7. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \\ -2 & 1 \end{bmatrix}$ and the vector $\mathsf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the matrix equation $\mathsf{A}\,\mathsf{x} = \mathsf{b}$.

(b) Verify that the vector $\mathsf{A}\,\hat{\mathsf{x}} - \mathsf{b}$, where $\hat{\mathsf{x}}$ is the least-squares solution found in part (7a), belongs to the space $R(\mathsf{A})^{\perp}$, the orthogonal complement of $R(\mathsf{A})$.

8. Suppose that a matrix $\mathsf{A} \in \mathbb{R}^{3,3}$ has eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 4$.

(a) Find the trace of A, find the trace of A^2, and find the determinant of A.

(b) Is matrix A invertible? If your answer is "yes", then prove it and find $\det(\mathsf{A}^{-1})$; if your answer is "no", then prove it.

9. Consider the matrix $\mathsf{A} = \begin{bmatrix} 7 & 5 \\ 3 & -7 \end{bmatrix}$.

(a) Find the eigenvalues and eigenvectors of A.

(b) Compute the matrix e^{A}.

10. Find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solution of the initial value problem

$$\frac{d}{dt}\mathsf{x}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where the matrix $\mathsf{A} = \begin{bmatrix} -5 & 2 \\ -12 & 5 \end{bmatrix}$ and the vector $\mathsf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Practice Exam 3. (Two hours.)

1. (a) Find the LU-factorization of the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & 2 \\ 2 & -8 & -3 \end{bmatrix}$.

(b) Use the LU-factorization above to find the solution of the linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$, where $\mathsf{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$.

2. Determine whether the following sets W_1 and W_2 are subspaces of the vector space $\mathbb{P}_2\big([0,1]\big)$. If your answer is "yes," find a basis of the subspace.

(a) $W_1 = \{\boldsymbol{p} \in \mathbb{P}_2\big([0,1]\big) \ : \ \displaystyle\int_0^1 \boldsymbol{p}(x)\,dx \leqslant 1 \ \}$;

(b) $W_2 = \{\boldsymbol{p} \in \mathbb{P}_2\big([0,1]\big) \ : \ \displaystyle\int_0^1 x\,\boldsymbol{p}(x)\,dx = 0 \ \}$.

3. Find a basis of $\mathbb{R}^4$ containing a basis of the null space of matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & -4 & 3 \\ 2 & 1 & 1 & 3 \\ 1 & 1 & -1 & 2 \\ 3 & 2 & 0 & 5 \end{bmatrix}.$$

4. Given a matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, introduce its characteristic polynomial $p_A(\lambda) = \det(\mathsf{A} - \lambda\,\mathsf{I}_n)$. This polynomial has the form $p_A(\lambda) = a_0 + a_1\lambda + \cdots + a_n\lambda^n$ for appropriate scalars $a_0, \cdots, a_n$. Now introduce a matrix-valued function $P_A : \mathbb{F}^{n,n} \to \mathbb{F}^{n,n}$ as follows

$$P_A(\mathsf{X}) = a_0\,\mathsf{I}_n + a_1\mathsf{X} + \cdots + a_n\mathsf{X}^n.$$

Determine whether the following statement is true or false and justify your answer: If matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ is diagonalizable, then $P_A(\mathsf{A}) = \mathsf{0}$. (That is, $P_A(\mathsf{A})$ is the zero matrix.)

5. Consider the vector space $\mathbb{R}^2$ with ordered bases

$$\mathcal{S} = \Big(\mathsf{e}_{1s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_s, \mathsf{e}_{2s} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_s\Big), \qquad \mathcal{U} = \Big(\mathsf{u}_{1s} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_s\Big).$$

Let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation given by

$$[\boldsymbol{T}(\mathsf{u}_1)]_s = \begin{bmatrix} -3 \\ 1 \end{bmatrix}_s, \quad [\boldsymbol{T}(\mathsf{u}_2)]_s = \begin{bmatrix} 1 \\ 3 \end{bmatrix}_s.$$

(a) Find the matrix T_{ss}.

(b) Find the matrix T_{uu}.

6. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix}$ and the vector $\mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the matrix equation $\mathsf{A}\,\mathsf{x} = \mathsf{b}$.

(b) Find the vector on $R(\mathsf{A})$ that is the closest to the vector b in $\mathbb{R}^3$.

7. Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$ and the subspace $W \subset \mathbb{C}^3$ given by

$$W = \mathrm{Span}\Big(\Big\{ \begin{bmatrix} i \\ 1 \\ i \end{bmatrix} \Big\}\Big).$$

(a) Find a basis for $W^\perp$, the orthogonal complement of W.

(b) Find an orthonormal basis for $W^\perp$.

8. (a) Find the eigenvalues and eigenvectors of matrix $\mathsf{A} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$, and show that A is diagonalizable.

(b) Knowing that matrix A above is diagonalizable, explicitly find a square root of matrix A, that is, find a matrix X such that $\mathsf{X}^2 = \mathsf{A}$. How many square roots does matrix A have?

9. Consider the matrix $\mathsf{A} = \begin{bmatrix} 8 & -18 \\ 3 & -7 \end{bmatrix}$.

(a) Find the eigenvalues and eigenvectors of A.

(b) Compute explicitly the matrix-valued function $e^{\mathsf{A}\,t}$ for $t \in \mathbb{R}$.

10. Find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solution of the initial value problem

$$\frac{d}{dt}\mathsf{x}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where the matrix $\mathsf{A} = \begin{bmatrix} -7 & 12 \\ -4 & 7 \end{bmatrix}$ and the vector $\mathsf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

10.3. Answers to exercises

Chapter 1: Linear systems

Section 1.1: Row and Column Pictures

1.1.1.- $x = 2$, $y = 3$.

1.1.2.- The system is not consistent.

1.1.3.- The nonlinear system has two solutions,

$$\big(x = \sqrt{2}, y = \sqrt{2}\big),$$
$$\big(x = -\sqrt{2}, y = -\sqrt{2}\big).$$

1.1.4.- The system is inconsistent.

1.1.5.- Subtract the first equation from the second. To the resulting equation subtract the third equation. One gets $0 = 1$. The system is inconsistent.

1.1.6.- $k = \pm 2$.

1.1.7.- The system is consistent. The solution is

$$x_1 = -1, \quad x_2 = 1.$$

1.1.8.-

(a) The vectors A_1 and A_2 are collinear, while vector b is not parallel to A_1 and A_2.

(b) The system is not consistent.

(c) The system is consistent and the solution is not unique. The solutions are $x_1 = 3 + x_2/2$ and x_2 arbitrary.

1.1.9.- $h = 1$.

1.1.10.- $\mathsf{A}_3 = -\mathsf{A}_1 + 2\mathsf{A}_2$. The system

$$\mathsf{A}_1 x_1 + \mathsf{A}_2 x_2 + \mathsf{A}_3 x_3 = \mathsf{0}$$

is consistent with infinitely many solutions given by

$$x_1 = x_3, \quad x_2 = -2x_3,$$

and x_3 arbitrary.

Section 1.2: Gauss-Jordan method

1.2.1.- $x_1 = 1, \quad x_2 = 0, \quad x_3 = 0.$

1.2.2.- $x_1 = 3, \quad x_2 = 1, \quad x_3 = 0.$

1.2.3.- $x_1 = 1, \quad x_2 = 0, \quad x_3 = -1.$

1.2.4.- $x_1 = 2, \quad x_2 = 4, \quad x_3 = 5.$
$x_1 = -4, \quad x_2 = -7, \quad x_3 = -8.$

1.2.5.- $x_1 = 1, \quad x_2 = 1, \quad x_3 = 1.$
$x_1 = 1, \quad x_2 = 2, \quad x_3 = 2.$
$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3.$

Section 1.3: Echelon forms

1.3.1.- $\text{rank}(\mathsf{A}) = 2$, pivot columns are the first and fourth.

1.3.2.- $x_1 = -x_4$, $x_2 = 0$, $x_3 = -2x_4$ and x_4 is a free variable.

1.3.3.- $x_1 = -1 + 2x_3$, $x_2 = 2 - 3x_3$ and x_3 is a free variable.

1.3.4.- For example:

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \mathsf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

1.3.5.- Matrix A has a reduced echelon form E_A with a pivot on every column, therefore E_A has no row with all coefficients zero. Hence, any system with such coefficient matrix A is always consistent.

1.3.6.-

(a) $k = -1$.

(b) $\mathsf{x} = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix}$.

Section 1.4: Non-homogeneous equations

1.4.1.- $\mathsf{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} x_4.$

1.4.2.- $\mathsf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix} x_3.$

1.4.3.- Hint: Recall that the matrix-vector product is a linear operation.

1.4.4.- $\mathsf{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} x_4 . + \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}.$

1.4.5.- $\mathsf{x} = \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix} x_3.$ Writing the solution as $\mathsf{x} = \mathsf{y} + \mathsf{z}x_3$, the solution vectors have end points on line passing through the end point of y and the line is tangent to the line parallel to z.

1.4.6.- $\mathsf{x} = \begin{bmatrix} 0 \\ 8 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ -5 \\ 1 \end{bmatrix} x_4.$ Writing the solution as $\mathsf{x} = \mathsf{y} + \mathsf{z}x_4$, the solution vectors have end points on line passing through the end point of y and the line is tangent to the line parallel to z.

1.4.7.-

(a) For $k \neq 3$ there exists a unique solution.

(b) If $k = 3$ the solutions are

$$\mathsf{x} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -3/2 \\ 1 \end{bmatrix} x_3.$$

Section 1.5: Floating-point numbers

1.5.1.-

(a) $x_1 = 0$, $x_2 = -1$.

(b) **Hint:** Modify the coefficients of the original system so that the approximate solution is an exact solution of the modified system.
One answer is: We keep the first equation coefficients and we modify the sources:

$$10^{-3} x_1 - x_2 = 1$$
$$x_1 + x_2 = -1.$$

(c) $x_1 = 1$, $x_2 = -1$.

(d) One answer is: We keep the first equation coefficients and we modify the sources:

$$10^{-3} x_1 - x_2 = 1 + 10^{-3}$$
$$x_1 + x_2 = 0.$$

(e) $x_1 = \dfrac{1}{1 + 10^{-3}}$, $x_2 = \dfrac{-1}{1 + 10^{-3}}$.

(f) $x_1 = 0.999$, $x_2 = -0.999$.

1.5.2.-

(a) $x_1 = 0$, $x_2 = 1$.

(b) $x_1 = 2$, $x_2 = 1$.

(c) $x_1 = 2$, $x_2 = 1$.

(d) $x_1 = 2$, $x_2 = 1$.

(e) $x_1 = \dfrac{2}{1 + 10^{-3}}$,

$$x_2 = \frac{1 + 3 \times 10^{-3}}{1 + 10^{-3}}.$$

1.5.3.- Without partial pivoting:

$$x_1 = 1.01, \quad x_2 = 1.03.$$

With partial pivoting:

$$x_1 = 1, \quad x_2 = 1.$$

Solution in $\mathbb{R}$:

$$x_1 = 1, \quad x_2 = 1.$$

Chapter 2: Matrix algebra

Section 2.1: Linear transformations

2.1.1.-
(a) T is not linear.
(b) T is linear.
(c) T is not linear.

2.1.2.- $\mathsf{x} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$.

2.1.3.- $\mathsf{x} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix} x_3$.

2.1.4.- $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 + x_2 \\ x_1 + 3x_2 \end{bmatrix}$.

2.1.5.-
(a) Rotation by an angle π.
(b) Expansion by 2.
(c) Projection onto x_2 axis.
(d) Reflection along the line $x_1 = x_2$.

2.1.6.- $\mathsf{A} = \begin{bmatrix} -2 & 7 \\ 5 & -3 \end{bmatrix}$.

Section 2.2: Linear combinations

2.2.1.-
(a) $\mathsf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
(b) $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 6 \end{bmatrix}$.
(c) $\mathsf{A} = \begin{bmatrix} 1 & 2+i & 3 \\ 2-i & 5 & 4-i \\ 3 & 4+i & 6 \end{bmatrix}$.

2.2.2.- $x = -1/2$, $y = -6$, $z = 0$.

2.2.3.- Hint: Compute the transpose of $\mathsf{A} + \mathsf{A}^T$. Do the same with $\mathsf{A} - \mathsf{A}^T$.

2.2.4.- Hint: Express $\mathsf{A} = \mathsf{C} + \mathsf{D}$, with C symmetric and D skew-symmetric. Find the expressions of matrices C and D in terms of matrix A.

2.2.5.-
(a) Hint: Introduce $i = j$ in the skew-symmetric matrix component condition.
(b) Hint: Introduce $i = j$ in the skew-Hermitian matrix component condition.
(c) Hint: Compute B^T.

2.2.6.- Hint: Use the properties of the adjoint (complex conjugate and transpose) operation.

2.2.7.- Hint: See the proof of Theorem 2.2.8.

Section 2.3: Matrix multiplication

2.3.1.-
(a) BA, CB not possible,

$$\mathsf{AB} = \begin{bmatrix} 10 & 15 \\ 12 & 8 \\ 28 & 52 \end{bmatrix}, \quad \mathsf{C}^T\mathsf{B} = \begin{bmatrix} 10 & 31 \end{bmatrix}.$$

(b) B^2 is not possible, and

$$\mathsf{A}^2 = \begin{bmatrix} 13 & -1 & 19 \\ 16 & 13 & 12 \\ 36 & -17 & 64 \end{bmatrix},$$

$$\mathsf{C}^T\mathsf{C} = \begin{bmatrix} 14 \end{bmatrix}, \quad \mathsf{C}\mathsf{C}^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

2.3.2.-
(a) $[\mathsf{A}, \mathsf{B}] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.
(b) $\mathsf{ABC} = \begin{bmatrix} 9 & 26 \\ 12 & 33 \end{bmatrix}$.

2.3.3.-

$$\mathsf{A}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathsf{A}^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

2.3.4.- $\mathsf{A}^n = \begin{bmatrix} 1 & na \\ 0 & 1 \end{bmatrix}$.

2.3.5.- Hint: Expand $(\mathsf{A}+\mathsf{B})^2$ and recall that $\mathsf{AB} = \mathsf{BA}$ iff $[\mathsf{A}, \mathsf{B}] = \mathsf{0}$.

2.3.6.- If $\mathsf{B} = a \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then

$$\mathsf{A} = \left[\begin{array}{ccc} \mathsf{I}_3 & \vdots & \mathsf{B} \\ \cdots & \cdots & \cdots \\ \mathsf{0} & \vdots & \mathsf{B} \end{array}\right].$$

Recalling that $a = 1/3$, we obtain,

$$\mathsf{A}^n = \left[\begin{array}{ccc} \mathsf{I}_3 & \vdots & 2^{(n-1)}\mathsf{B} \\ \cdots & \cdots & \cdots \\ \mathsf{0} & \vdots & \mathsf{B} \end{array}\right].$$

2.3.7.- Hint: Write down in components each side of the equation.

2.3.8.- Hint: Express $\mathrm{tr}\,(\mathsf{A}^T\mathsf{A})$ in components, and show that it is the sum of squares of every component of matrix A.

2.3.9.- Hint: Recall that AB is symmetric iff $(\mathsf{AB})^T = \mathsf{AB}$.

2.3.10.- Hint: Take the trace on each side of $[\mathsf{A}, \mathsf{X}] = \mathsf{I}$ and use the properties of the trace to get a contradiction.

Section 2.4: Inverse matrix

2.4.1.-

(a) $\mathsf{A}^{-1} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}$.

(b) $\mathsf{A}^{-1} = \begin{bmatrix} -2 & 3 & -10 \\ 0 & -1 & 3 \\ -1 & 1 & -4 \end{bmatrix}$.

(c) A is not invertible.

2.4.2.- $k = 2$ or $k = -3$.

2.4.3.-

(a) $\mathsf{A}^{-1} = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ -1 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix}$.

(b) $\mathsf{x} = \begin{bmatrix} 3/2 \\ -1 \\ -4 \end{bmatrix}$.

2.4.4.- Hint: Just write down the commutator of two matrices:

$$[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA},$$

with $\mathsf{B} = \mathsf{A}^{-1}$.

2.4.5.- $\mathsf{X} = \begin{bmatrix} 2 & 4 \\ -1 & -2 \\ 3 & 3 \end{bmatrix}$.

2.4.6.- Hint: Use that

$$\left(\mathsf{A}^{-1}\right)^T = \left(\mathsf{A}^T\right)^{-1}.$$

2.4.7.- Hint: Generalize for matrices the identity for real numbers,

$$(1-a)(1+a) = 1 - a^2.$$

2.4.8.- Hint: Generalize for matrices the identity for real numbers,

$$(1-a)(1+a+a^2) = 1 - a^3.$$

2.4.9.-

(a) Hint: Use Theorem 2.4.3.

(b) Hint: Use Theorem 2.4.3.

(c) Hint: $\text{tr}\,(\mathsf{AB}) = \text{tr}\,(\mathsf{BA})$.

2.4.10.- Hint: Multiply the matrices using block multiplication.

Section 2.5: Null and Range spaces

2.5.1.-

$$R(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \Big\}\Big),$$

$$R(\mathsf{A}^T) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix} \Big\}\Big).$$

2.5.2.-

$$N(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -4 \\ 0 \\ 1 \end{bmatrix} \Big\}\Big),$$

$$R(\mathsf{A}) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \Big\}\Big),$$

$$N(\mathsf{A}^T) = \text{Span}\Big(\Big\{ \begin{bmatrix} -2 \\ -1/2 \\ 1 \end{bmatrix} \Big\}\Big),$$

$$R(\mathsf{A}^T) = \text{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ 0 \\ 4 \\ -2 \end{bmatrix} \Big\}\Big).$$

2.5.3.-

(a) The system $\mathsf{Ax} = \mathsf{b}$ is consistent since $\mathsf{b} \in R(\mathsf{A})$, because of

$$\mathsf{b} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \in R(\mathsf{A}).$$

(b) Since $N(\mathsf{A}) \neq \{\boldsymbol{0}\}$, given any solution x of the system $\mathsf{Ax} = \mathsf{b}$, then another solution is

$$\tilde{\mathsf{x}} = \mathsf{x} + c \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix},$$

for any $c \in \mathbb{R}$.

2.5.4.-

(a) A linear system $\mathsf{Ax} = \mathsf{b}$ is consistent iff b is a linear combination of the columns of A iff $\mathsf{b} \in R(\mathsf{A})$.

(b) Hint: $\mathsf{A}(\mathsf{x}+\hat{\mathsf{x}}) = \mathsf{Ax}$ for every vector $\hat{\mathsf{x}} \in N(\mathsf{A})$.

2.5.5.- Theorem 2.4.3 part (c) says that A is invertible iff $\mathsf{E}_\mathsf{A} = \mathsf{I}$, which is equivalent to say that $R(\mathsf{A}) = \mathbb{F}^n$.

2.5.6.-

$$N(\mathsf{A}) = N(\mathsf{A}^T) = \{\boldsymbol{0}\},$$
$$R(\mathsf{A}) = R(\mathsf{A}^T) = \mathbb{F}^n.$$

2.5.7.-

(a) No, since $\mathsf{E}_{\mathsf{A}^T} \neq \mathsf{E}_{\mathsf{B}^T}$.

(b) Yes, since $\mathsf{E}_\mathsf{A} = \mathsf{E}_\mathsf{B}$.

(c) Yes, since $\mathsf{E}_\mathsf{A} = \mathsf{E}_\mathsf{B}$.

(d) No, since $\mathsf{E}_{\mathsf{A}^T} \neq \mathsf{E}_{\mathsf{B}^T}$.

Section 2.6: LU-factorization

2.6.1.- $\mathsf{A} = \mathsf{LU}$ where

$$\mathsf{L} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \ \mathsf{U} = \begin{bmatrix} 5 & 2 \\ 0 & 3 \end{bmatrix}.$$

2.6.2.- $\mathsf{A} = \mathsf{LU}$ where

$$\mathsf{L} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}, \ \mathsf{U} = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 4 & 1 \end{bmatrix}.$$

2.6.3.- $\mathsf{A} = \mathsf{LU}$ where

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -2 & 1 \end{bmatrix}, \ \mathsf{U} = \begin{bmatrix} 2 & 1 & 2 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

2.6.4.- Matrix A does not hav an LU-factorization.

2.6.5.- $\mathsf{T} = \mathsf{LU}$ where

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1/2 & 1 & 0 & 0 \\ 0 & -2/3 & 1 & 0 \\ 0 & 0 & -3/4 & 1 \end{bmatrix},$$

$$\mathsf{U} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 3/2 & -1 & 0 \\ 0 & 0 & 4/3 & -1 \\ 0 & 0 & 0 & 1/4 \end{bmatrix}.$$

2.6.6.- $\mathsf{A} = \mathsf{LU}$ where

$$\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}, \ \mathsf{U} = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 4 \end{bmatrix}.$$

Then,

$$\mathsf{y} = \begin{bmatrix} 12 \\ 0 \\ -24 \end{bmatrix}, \ \mathsf{x} = \begin{bmatrix} 6 \\ 6 \\ -6 \end{bmatrix}.$$

2.6.7.- $c = 0$ or $c = \pm\sqrt{2}$.

Chapter 3: Determinants

Section 3.1: Definitions and properties

3.1.1.- $\det(\mathsf{A}) = 8$ and $\det(\mathsf{B}) = -8$.

3.1.2.- $V = 14$.

3.1.3.-

$$\det(\mathsf{U}) = (1)(4)(6) = 24,$$
$$\det(\mathsf{L}) = (1)(3)(6) = 18.$$

3.1.4.- $\det(\mathsf{A}) = -8$.

3.1.5.- For example:

$$\mathsf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathsf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

On the one hand $\det(\mathsf{A} + \mathsf{B}) = 1$, while on the other hand $\det(\mathsf{A}) = \det(\mathsf{B}) = 0$.

3.1.6.-

$$\det(2\mathsf{A}) = 2^n \det(\mathsf{A})$$
$$\det(-\mathsf{A}) = (-1)^n \det(\mathsf{A})$$
$$\det(\mathsf{A}^2) = \big[\det(\mathsf{A})\big]^2.$$

3.1.7.-

$$\begin{aligned} \det(\mathsf{B}) &= \det(\mathsf{P}^{-1}\mathsf{A}\mathsf{P}) \\ &= \det(\mathsf{P}^{-1}) \det(\mathsf{A}) \det(\mathsf{P}) \\ &= \frac{1}{\det(\mathsf{P})} \det(\mathsf{A}) \det(\mathsf{P}) \\ &= \det(\mathsf{A}). \end{aligned}$$

3.1.8.-

$$\begin{aligned} \det(\mathsf{A}^*) &= \det\big(\overline{\mathsf{A}}^T\big) \\ &= \det\big(\overline{\mathsf{A}}\big) \\ &= \overline{\det(\mathsf{A})}. \end{aligned}$$

3.1.9.-

$$\begin{aligned} \det\big(\mathsf{A}^*\mathsf{A}\big) &= \det\big(\mathsf{A}^*\big) \det(\mathsf{A}) \\ &= \overline{\det(\mathsf{A})} \det(\mathsf{A}) \\ &= \big|\det(\mathsf{A})\big|^2 \geqslant 0. \end{aligned}$$

3.1.10.-

$$\begin{aligned} \det(k\mathsf{A}) &= \big|k\mathsf{A}_1, \cdots, k\mathsf{A}_n\big| \\ &= k^n \big|\mathsf{A}_1, \cdots, \mathsf{A}_n\big| \\ &= k^n \det(\mathsf{A}). \end{aligned}$$

3.1.11.-

$$\begin{aligned} \det(\mathsf{A}) &= \det\big(\mathsf{A}^T\big) \\ &= \det(-\mathsf{A}) \\ &= (-1)^n \det(\mathsf{A}) \\ &= -\det(\mathsf{A}) \quad \Rightarrow \quad 2\det(\mathsf{A}) = 0. \end{aligned}$$

Therefore, $\det(\mathsf{A}) = 0$.

3.1.12.-

$$\begin{aligned} 1 &= \det(\mathsf{I}_n) \\ &= \det\big(\mathsf{A}^T\mathsf{A}\big) \\ &= \det\big(\mathsf{A}^T\big) \det(\mathsf{A}) \\ &= \big[\det(\mathsf{A})\big]^2 \quad \Rightarrow \quad \det(\mathsf{A}) = \pm 1. \end{aligned}$$

Section 3.2: Applications

3.2.1.- $\mathsf{A}^{-1} = \dfrac{1}{8}\begin{bmatrix} 0 & 1 & -1 \\ -8 & 4 & 4 \\ 16 & -6 & -2 \end{bmatrix}$.

3.2.2.-

$$\big(\mathsf{A}^{-1}\big)_{12} = \frac{8}{41}, \quad \big(\mathsf{A}^{-1}\big)_{32} = -\frac{19}{41}.$$

3.2.3.- $\det(\mathsf{A}) = 10$ and $\det(\mathsf{B}) = 0$.

3.2.4.- Hint:

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = \begin{vmatrix} 1 & a & a^2 \\ 0 & (b-a) & (b^2-a^2) \\ 0 & 0 & (c-a)(c-b) \end{vmatrix}.$$

3.2.5.- $k \neq \pm 1$.

3.2.6.- $\mathsf{x} = \dfrac{1}{(ad-bc)}\begin{bmatrix} d \\ -c \end{bmatrix}$.

3.2.7.- $\mathsf{x} = \begin{bmatrix} -2 \\ 4 \\ -1 \end{bmatrix}$.

Chapter 4: Vector Spaces

Section 4.1: Spaces and subspaces

4.1.1.-

(a) No.
(b) Yes.
(c) No.
(d) Yes.
(e) No.
(f) No.

4.1.2.-

(a) Yes.
(b) No.
(c) No.
(d) Yes.
(e) No.
(f) Yes.
(g) Yes.

4.1.3.- The result is $W_1+W_2=\mathbb{R}^3$. An example is $W_1 = \text{Span}\big(\{\mathtt{e}_1,\mathtt{e}_2\}\big)$, and $W_2 = \text{Span}\big(\{\mathtt{e}_3\}\big)$. Then,

$$W_1 + W_2 = \text{Span}\big(\{\mathtt{e}_1,\mathtt{e}_2,\mathtt{e}_3\}\big) = \mathbb{R}^3.$$

4.1.4.-

(a) The line $\text{Span}\Big(\Big\{\begin{bmatrix}1\\2\\3\end{bmatrix}\Big\}\Big)$.

(b) The plane $\text{Span}\Big(\Big\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\Big\}\Big)$.

(c) The space $\mathbb{R}^3$.

4.1.5.- If $S_1=\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_k\}$ and $S_2=\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_l\}$, then

$$\text{Span}(S_1\cup S_2) =$$

$$\{(a_1\boldsymbol{u}_1+\cdots+a_k\boldsymbol{u}_k)+(b_1\boldsymbol{v}_1+\cdots+b_l\boldsymbol{v}_l)\}$$

for every set of scalars $a_1,\cdots,a_k\in\mathbb{F}$, $b_1\cdots,b_l\in\mathbb{F}$. From the definition of sum of subspaces we see that this last subspace can be rewritten as

$$\text{Span}(S_1\cup S_2) =$$

$$\{a_1\boldsymbol{u}_1+\cdots+a_k\boldsymbol{u}_k\}+\{b_1\boldsymbol{v}_1+\cdots+b_l\boldsymbol{v}_l\}$$
$$= \text{Span}(S_1)+\text{Span}(S_2).$$

4.1.6.- Find a vector not in $\text{Span}(W_1)$. Denoting $\boldsymbol{u}=\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}$, we compute

$$\begin{bmatrix}1&1&u_1\\2&0&u_2\\3&1&u_3\end{bmatrix}\to$$

$$\to\begin{bmatrix}1&1&u_1\\0&-2&u_2-2u_1\\0&0&u_3-u_1-u_2\end{bmatrix}.$$

Choose any solution of $u_3-u_1-u_2\neq 0$. For example, $\boldsymbol{u}=\begin{bmatrix}1\\0\\0\end{bmatrix}$, and

$$W_2=\text{Span}(\{\boldsymbol{u}\}).$$

Section 4.2: Linear dependence

4.2.1.-

(a) Linearly dependent.

(b) Linearly dependent. (A four vector set in $\mathbb{R}^3$ is always linearly dependent.)

(c) Linearly independent.

4.2.2.-

(a) One possible solution is:

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \right\}.$$

(b) 11 sets.

4.2.3.- If $S = \{\boldsymbol{0}, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$, then

$$1\boldsymbol{0} + 0\boldsymbol{v}_1 + \cdots + 0\boldsymbol{v}_k = \boldsymbol{0}.$$

4.2.4.- Hint: Show that the constants c_1, c_2 sastisfying

$$c_1(\boldsymbol{v} + \boldsymbol{w}) + c_2(\boldsymbol{v} - \boldsymbol{w}) = \boldsymbol{0}$$

are both zero iff the constants $\tilde{c}_1$, $\tilde{c}_2$ satisfying

$$\tilde{c}_1\boldsymbol{v} + \tilde{c}_2\boldsymbol{w} = \boldsymbol{0}$$

are both zero. Once that is proven, and since the latter statement is true, so is the former.

4.2.5.- Linearly dependent.

4.2.6.- Hint: Use the definition. Prove that there are non-zero constants c_1, c_2, c_3, c_4 solutions of

$$c_1 + c_2 c + c_3 x^2 + c_4(1 + x + x^2) = 0.$$

4.2.7.- Linearly independent.

Section 4.3: Bases and dimension

4.3.1.- Bases are not unique. Here are possible choices:

$$\mathcal{N}_A = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\},$$

$$\mathcal{R}_A = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \right\},$$

$$\mathcal{N}_{A^T} = \left\{ \begin{bmatrix} 1/3 \\ -5/3 \\ 1 \end{bmatrix} \right\},$$

$$\mathcal{R}_{A^T} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \\ 3 \end{bmatrix} \right\}.$$

4.3.2.- The dimension is 3.

4.3.3.-

(a) $\dim \mathbb{P}_n = n + 1$.
(b) $\dim \mathbb{F}^{m,n} = mn$.
(c) $\dim(\text{Sym}_n) = n(n+1)/2$.
(d) $\dim(\text{SkewSym}_n) = n(n-1)/2$.

4.3.4.- One example is the following: $\mathsf{v}_1 = \mathsf{e}_1$, $\mathsf{v}_2 = \mathsf{e}_2$, and

$$W = \text{Span}\left(\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \right).$$

4.3.5.- To verify that $\mathsf{v} \in N(\mathsf{A})$ compute $\mathsf{A}\mathsf{v}$ and check that the result is $\mathsf{0}$. A basis for $N(\mathsf{A})$ containing v is

$$\mathcal{N} = \left\{ \begin{bmatrix} -8 \\ 1 \\ 3 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

4.3.6.- The set $\mathcal{B}$ is a basis of that subspace.

Section 4.4: Vector components

4.4.1.-

$$\mathsf{v}_u = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

4.4.2.-

$$\mathsf{v}_u = \begin{bmatrix} 9 \\ 2 \\ -3 \end{bmatrix}.$$

4.4.3.-

(a) $\mathsf{r}_s = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$.

(b) $\mathsf{r}_q = \begin{bmatrix} 6 \\ -4 \\ -1 \end{bmatrix}$.

4.4.4.-

(a) Hint: Prove that $\text{Span}(\mathcal{M}) = \mathbb{R}^{2,2}$, and that $\mathcal{M}$ is linearly independent. For the first part, show that for every matrix

$$\mathsf{B} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

there exist constants c_1, c_2, c_3, c_4 solution of

$$c_1\mathsf{M}_1 + c_2\mathsf{M}_2 + c_3\mathsf{M}_3 + c_4\mathsf{M}_4 = \mathsf{B}.$$

For the second part, choose $\mathsf{B} = \mathsf{0}$ and show that the only constants c_1, c_2, c_3, c_4 solution of

$$c_1\mathsf{M}_1 + c_2\mathsf{M}_2 + c_3\mathsf{M}_3 + c_4\mathsf{M}_4 = \mathsf{0}$$

are $c_1 = c_2 = c_3 = c_4 = 0$.

(b)

$$\mathsf{A} = \frac{5}{2}\mathsf{M}_1 + \frac{1}{2}\mathsf{M}_2 + \frac{5}{2}\mathsf{M}_3 - \frac{3}{2}\mathsf{M}_4.$$

Using the notation

$$\mathsf{D}_M = \begin{bmatrix} d_1 & d_2 \\ d_3 & d_4 \end{bmatrix} \quad \Leftrightarrow$$

$$\mathsf{D} = d_1\mathsf{M}_1 + d_2\mathsf{M}_2 + d_3\mathsf{M}_3 + d_4\mathsf{M}_4$$

then we get

$$\mathsf{A}_M = \frac{1}{2}\begin{bmatrix} 5 & 1 \\ 5 & -3 \end{bmatrix}.$$

Chapter 5: Linear transformations

Section 5.1: Linear transformations

5.1.1.- The projection $\boldsymbol{A}$ neither injective nor surjective.

5.1.2.- The rotation $\mathsf{R}(\theta)$ is both injective and surjective for $\theta \in [0, 2\pi)$.

5.1.3.-

(a) Linear.
(b) Linear.
(c) Non-linear.
(d) Linear.
(e) Linear.

5.1.4.- Hint: Show that

$$\boldsymbol{T}(a\mathsf{x} + b\mathsf{y}) = a\,\boldsymbol{T}(\mathsf{x}) + b\,\boldsymbol{T}(\mathsf{y})$$

for all $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ and all $a, b \in \mathbb{F}$.
$\boldsymbol{T}$ is not a linear operator, it is a linear functional.

5.1.5.- Hint: Show that

$$\Delta(a\boldsymbol{p} + b\boldsymbol{q}) = a\Delta(\boldsymbol{p}) + b\Delta(\boldsymbol{q})$$

for all $\boldsymbol{p}, \boldsymbol{p} \in \mathbb{P}_3$ and all $a, b \in \mathbb{F}$.
Δ is not injective, but it is surjective.

5.1.6.- Hint: Recall the definition of a linearly independent set.

5.1.7.- Hint: Recall the Nullity-Rank Theorem 5.1.6.

5.1.8.- Hint: Recall the Nullity-Rank Theorem 5.1.6.

Section 5.2: The inverse transformations

5.2.1.- Hint: Show that T is bijective. The inverse transformation is

$$\boldsymbol{T}^{-1}\Big(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\Big) = \frac{1}{5}\begin{bmatrix} 2y_1 + y_2 \\ 3y_1 - y_2 \end{bmatrix}.$$

5.2.2.- Hint: Show that Δ is bijective. The inverse transformation is

$$\Delta^{-1}(b_0 + b_1 x + b_2 x^2) = \frac{b_0}{2}x^2 + \frac{b_1}{6}x^3 + \frac{b_2}{12}x^4.$$

5.2.3.- Hint: Choose any isomorphism $T : V \to W$ and use the Nullity-Rank Theorem.

5.2.4.- Hint: Use the coordinate map to find an isomorphism between these spaces.

5.2.5.- Hint: Generalize the proof in Exercise **5.2.4**.

5.2.6.- Hint: Generalize the proof in Example 5.2.8.

5.2.7.- Hint: Use the ideas given in Example 5.2.9.

5.2.8.- The basis $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}$ is given by

$$\boldsymbol{\phi}_1\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = -x_1 + 3x_2,$$

$$\boldsymbol{\phi}_2\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = x_1 - 2x_2.$$

Using row vector notation,

$$[\boldsymbol{\phi}_1] = \begin{bmatrix} -1, 3 \end{bmatrix},$$
$$[\boldsymbol{\phi}_2] = \begin{bmatrix} 1, -2 \end{bmatrix}.$$

Section 5.3: The algebra of linear operators

5.3.1.-

(a) $(3\boldsymbol{T}-2\boldsymbol{S})\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}3x_2\\x_1\end{bmatrix}$.

(b) $(\boldsymbol{T}\circ\boldsymbol{S})\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}x_1\\0\end{bmatrix}$.

$(\boldsymbol{S}\circ\boldsymbol{T})\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}0\\x_2\end{bmatrix}$.

(c) $\boldsymbol{T}^2\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}x_1\\x_2\end{bmatrix}$.

$\boldsymbol{S}^2\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}0\\0\end{bmatrix}$.

5.3.2.- $\dim L(\mathbb{F}^4)=16$, $\dim L(\mathbb{P}_2)=9$, $\dim L(\mathbb{F}^{3,2})=36$.

5.3.3.-

$$\boldsymbol{T}^{-2}\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}x_1/4\\-x_1+x_2\end{bmatrix}.$$

5.3.4.-

$$p(\boldsymbol{T})\Big(\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big)=\begin{bmatrix}2x_1+6x_2\\9x_1+11x_2\end{bmatrix}.$$

5.3.5.- Hint: Use that $\cos^2(\alpha)+\sin^2(\alpha)=1$.

5.3.6.- Hint: Use the definition of the commutator.

Section 5.4: Transformation components

5.4.1.- $\mathsf{T}_{uu}=\begin{bmatrix}2&0\\0&3\end{bmatrix}$.

5.4.2.-

(a) $\mathsf{T}_{uu}=\dfrac{1}{2}\begin{bmatrix}2&-3&1\\-2&1&1\\0&1&-1\end{bmatrix}$.

(b) To verify $[\boldsymbol{T}(\boldsymbol{v})]_u=\mathsf{T}_{uu}\mathsf{v}_u$ one must compute the left-hand side and the right-hand side, and check one obtains the same result. For the left-hand side:

$$[\boldsymbol{T}(\mathsf{v}_s)]_s=\begin{bmatrix}0\\0\\-1\end{bmatrix}_s$$

and

$$\begin{bmatrix}0\\0\\-1\end{bmatrix}_s=\frac{1}{2}\begin{bmatrix}-1\\-1\\1\end{bmatrix}_u.$$

For the right-hand side:

$$\mathsf{v}_u=\begin{bmatrix}1\\1\\0\end{bmatrix}_u,$$

and $\mathsf{T}_{uu}\mathsf{v}_u$ is given by

$$\frac{1}{2}\begin{bmatrix}2&-3&1\\-2&1&1\\0&1&-1\end{bmatrix}\begin{bmatrix}1\\1\\0\end{bmatrix}_u=\frac{1}{2}\begin{bmatrix}-1\\-1\\1\end{bmatrix}_u.$$

5.4.3.- Hint: Recall the definition

$$\mathsf{T}_{s\tilde{s}}=\big[[\boldsymbol{T}(\boldsymbol{e}_1)]_{\tilde{s}},\cdots[\boldsymbol{T}(\boldsymbol{e}_n)]_{\tilde{s}}\big].$$

If $\mathsf{A}=[\mathsf{A}_1,\cdots,\mathsf{A}_n]$, then for $i=1,\cdots n$ holds $[\boldsymbol{T}(\boldsymbol{e}_i)]_{\tilde{s}}=\mathsf{A}\mathsf{e}_i=\mathsf{A}_i$. Therefore,

$$\mathsf{T}_{s\tilde{s}}=\mathsf{A}.$$

5.4.4.- $\mathsf{T}_{s\tilde{s}}=\begin{bmatrix}0&-1&2&0\\0&0&-2&6\\0&0&0&-3\end{bmatrix}$.

5.4.5.-

$$(\mathsf{D}\circ\mathsf{S})_{s\tilde{s}}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}.$$

$$(\mathsf{S}\circ\mathsf{D})_{\tilde{s}s}=\begin{bmatrix}0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}.$$

Section 5.5: Change of basis

5.5.1.-

(a) $\mathsf{I}_{us} = \begin{bmatrix} 2 & 1 \\ -9 & 8 \end{bmatrix}$,

$\mathsf{I}_{su} = \dfrac{1}{25}\begin{bmatrix} 8 & -1 \\ 9 & 2 \end{bmatrix}$.

(b) $\mathsf{x}_s = \begin{bmatrix} 5 \\ -10 \end{bmatrix}$, $\mathsf{x}_u = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

5.5.2.-

(a) $\mathsf{I}_{bc} = \dfrac{1}{9}\begin{bmatrix} -1 & -6 & -2 \\ 2 & 3 & 4 \\ -5 & -3 & -1 \end{bmatrix}$,

$\mathsf{I}_{cb} = \begin{bmatrix} 1 & 0 & -2 \\ -2 & -1 & 0 \\ 1 & 3 & 1 \end{bmatrix}$.

(b) $\mathsf{x}_b = \begin{bmatrix} -3 \\ 0 \\ -3 \end{bmatrix}$, $\mathsf{x}_c = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$.

5.5.3.-

(a) $\mathsf{x}_s = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$.

(b) $\mathsf{e}_{1u} = \dfrac{1}{3}\begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\mathsf{e}_{2u} = \dfrac{1}{3}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

5.5.4.-

(a) $\mathsf{x}_s = \begin{bmatrix} -1 \\ 5 \end{bmatrix}$.

(b) $\mathsf{x}_b = \dfrac{1}{4}\begin{bmatrix} 3 \\ -7 \end{bmatrix}$.

5.5.5.- $\mathsf{b}_{1s} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$, $\mathsf{b}_{2s} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

5.5.6.- Hint: Find an invertible matrix P such that $\mathsf{C} = \mathsf{P}^{-1}\mathsf{A}\mathsf{P}$.

5.5.7.-

$$\mathsf{T}_{ss} = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 7 \end{bmatrix},$$

$$\mathsf{T}_{uu} = \begin{bmatrix} 1 & 4 & 3 \\ -1 & -2 & -9 \\ 1 & 1 & 8 \end{bmatrix}.$$

5.5.8.-

$$\mathsf{T}_{us} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix},\ \mathsf{T}_{ss} = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix},$$

$$\mathsf{T}_{uu} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix},\ \mathsf{T}_{su} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}.$$

Chapter 6: Inner product spaces

Section 6.1: The dot product

6.1.1.- $\|\mathsf{u}\| = 5$, $\|\mathsf{v}\| = 2$, $d(\mathsf{u},\mathsf{v}) = \sqrt{31}$, $\theta = \arccos(-1/10)$.

6.1.2.-

$$\mathsf{u} = \frac{1}{\sqrt{13}}\begin{bmatrix}-2\\3\end{bmatrix},\ \mathsf{v} = \frac{1}{\sqrt{13}}\begin{bmatrix}2\\-3\end{bmatrix}.$$

6.1.3.- $\mathsf{u} = \dfrac{1}{\sqrt{10}}\begin{bmatrix}1+2i\\2-i\end{bmatrix}$.

6.1.4.-

(a) $\left\{\begin{bmatrix}1\\0\end{bmatrix},\ \begin{bmatrix}1\\1\end{bmatrix}\right\}$.

(b) $\left\{\begin{bmatrix}0\\0\end{bmatrix},\ \begin{bmatrix}1\\1\end{bmatrix}\right\}$.

6.1.5.- Hint: Expand the expression

$$\|\mathsf{x}-\mathsf{y}\|^2 = (\mathsf{x}-\mathsf{x})\cdot\overline{(\mathsf{x}-\mathsf{y})}.$$

6.1.6.-

(a) Hint: For every $\mathsf{x},\mathsf{y}\in\mathbb{R}^n$ holds

$$\mathsf{Re}(\mathsf{x}\cdot\mathsf{y}) = 0 \Leftrightarrow \mathsf{x}\cdot\mathsf{y} = 0.$$

(b) Hint: For $\mathsf{x},\mathsf{y}\in\mathbb{C}^n$ satisfying $\mathsf{Im}(\mathsf{x}\cdot\mathsf{y})\neq 0$ holds

$$\mathsf{Re}(\mathsf{x}\cdot\mathsf{y}) = 0 \not\Rightarrow \mathsf{x}\cdot\mathsf{y} = 0.$$

6.1.7.- Hint: Use Problem **6.1.5**.

Section 6.2: Inner product

6.2.1.-

(a) No.
(b) No.
(c) Yes.
(d) No.

6.2.2.-

(a) Hint: Choose a particular $\boldsymbol{x}$.
(b) Hint: Use the properties of the inner product.

6.2.3.-

(a) Yes.
(b) No.
(c) No.

6.2.4.- Hint: $N(\mathsf{A}) = \{\mathsf{0}\}$ is needed to show the positivity property of the inner product.

6.2.5.- $k = -3$.

6.2.6.- $\|\mathsf{A}\|_F = \sqrt{10}$, $\|\mathsf{B}\|_F = \sqrt{3}$.

6.2.7.- Hint: Use properties of the trace operation.

6.2.8.- $\boldsymbol{q} = \pm\dfrac{1}{2}(3-5x^2)$.

Section 6.3: Orthogonal vectors

6.3.1.- Hint: ($\Leftarrow$): choose particular values for a, b. Choose $b = 1$, $a = \langle \mathsf{x}, \mathsf{y} \rangle$.

6.3.2.- $\mathsf{x} \in \text{Span}\Big(\Big\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \Big\}\Big)$.

6.3.3.- $\mathsf{x} \in \text{Span}\Big(\Big\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \Big\}\Big)$.

6.3.4.- Hint: Compute all inner products $\langle \boldsymbol{p}_i, \boldsymbol{p}_j \rangle$ for $i, j = 0, 1, 2, 3$.

6.3.5.-

(a) Show that $\mathsf{U}^T\mathsf{U} = \mathsf{I}_3$, where the columns in U are the basis vectors in $\mathcal{U}$.

(b) $[\mathsf{x}]_u = \mathsf{U}\mathsf{x} = \dfrac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} \\ -\sqrt{2} \\ -5 \end{bmatrix}$.

6.3.6.-

(a) Hint: Compute all inner products $\langle \mathsf{E}_i, \mathsf{E}_j \rangle$ for $i, j = 1, 2, 3, 4$.

(b) Hint: Use the definition, $[\mathsf{A}]_i = \langle \mathsf{M}_i, \mathsf{A} \rangle_F$.

6.3.7.- $\theta = \pi/2$.

6.3.8.- $\mathsf{U}_{:3} = \dfrac{1}{\sqrt{42}} \begin{bmatrix} -5 \\ 4 \\ 1 \end{bmatrix}$.

Section 6.4: Orthogonal projections

6.4.1.- $\mathsf{x} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

6.4.2.-

(a) $\mathsf{w}_2 = \begin{bmatrix} 1/2 \\ 1 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 3/2 \\ -2 \\ 5/2 \end{bmatrix}$.

(b) $\mathsf{x} = \dfrac{5}{3} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \dfrac{1}{3} \begin{bmatrix} 7 \\ -1 \\ -5 \end{bmatrix}$.

6.4.3.- $\mathsf{x} = \dfrac{1}{3} \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} + \dfrac{1}{3} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

6.4.4.- $\mathcal{R}^\perp = \Big\{ \begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix} \Big\}$.

6.4.5.-

(a) $\mathcal{W}^\perp = \Big\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \Big\}$.

(b) $\tilde{\mathcal{W}}^\perp = \Big\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \Big\}$.

6.4.6.- $\mathcal{W}^\perp = \Big\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Big\}$.

6.4.7.-

(a) Hint: Use the definition of the orthogonal complement of a subspace.

(b) Hint: Show $(X + Y)^\perp \subset X^\perp \cap Y^\perp$ using (a). Show the other inclusion, $X^\perp \cap Y^\perp \subset (X+Y)^\perp$ using the definition of orhtogonal complement of a subspace.

(c) Hint: Use part (b), that is, $(\tilde{X} + \tilde{Y})^\perp = \tilde{X}^\perp \cap \tilde{Y}^\perp$, for $\tilde{X} = X^\perp$ and $\tilde{Y} = Y^\perp$.

Section 6.5: Gram-Schmidt method

6.5.1.- $\left\{\frac{1}{3}\begin{bmatrix}-2\\2\\-1\end{bmatrix},\ \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\-1\\0\end{bmatrix}\right\}$.

6.5.2.-

(a) $\left\{\begin{bmatrix}0\\1\\0\end{bmatrix},\ \frac{1}{5}\begin{bmatrix}3\\0\\4\end{bmatrix}\right\}$.

(b) $\mathsf{x} = \frac{1}{5}\begin{bmatrix}9\\5\\12\end{bmatrix} + \frac{4}{5}\begin{bmatrix}4\\0\\-3\end{bmatrix}$.

6.5.3.-

$\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix},\ \frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\-1\end{bmatrix},\ \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\-1\end{bmatrix}\right\}$.

6.5.4.- $\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix},\ \frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\-1\end{bmatrix}\right\}$.

6.5.5.- $\left\{\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = -\frac{1}{2} + x,\ \boldsymbol{q}_2 = \frac{1}{6} - x + x^2\right\}$.

Section 6.6: The adjoint operator

6.6.1.-

6.6.2.-

Chapter 7: Approximation methods

Section 7.1: Best approximation

7.1.1.- **7.1.2.-**

Section 7.2: Least squares

7.2.1.- **7.2.2.-**

Section 7.3: Finite difference method

7.3.1.- **7.3.2.-**

Section 7.4: Finite element method

7.4.1.- **7.4.2.-**

10.4. Solutions to Practice Exams

Solutions to Practice Exam 1.

1. Consider the matrix $\mathsf{A} = \begin{bmatrix} 5 & 3 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix}$. Find the coefficients $(\mathsf{A}^{-1})_{21}$ and $(\mathsf{A}^{-1})_{32}$ of the matrix A^{-1}, that is, of the inverse matrix of A. Show your work.

SOLUTION: The formula for the inverse matrix $\mathsf{A}^{-1} = \mathsf{C}^T / \det(\mathsf{A})$ implies that

$$\left(\mathsf{A}^{-1}\right)_{21} = \frac{C_{12}}{\det(\mathsf{A})} \qquad \left(\mathsf{A}^{-1}\right)_{32} = \frac{C_{23}}{\det(\mathsf{A})}.$$

We start computing the $\det(\mathsf{A})$, that is,

$$\begin{vmatrix} 5 & 3 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{vmatrix} = 5\begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} - 3\begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = 15 - 9 - 3 \quad \Rightarrow \quad \det(\mathsf{A}) = 3.$$

Now the cofactors,

$$C_{12} = (-1)^{(1+2)}\begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} = -3, \qquad C_{23} = (-1)^{(2+3)}\begin{vmatrix} 5 & 3 \\ 2 & 1 \end{vmatrix} = 1.$$

Therefore, we conclude that

$$\boxed{\left(\mathsf{A}^{-1}\right)_{21} = -1}, \qquad \boxed{\left(\mathsf{A}^{-1}\right)_{32} = \frac{1}{3}}.$$

◁

2. (a) Find $k \in \mathbb{R}$ such that the volume of the parallelepiped formed by the vectors below is equal to 4, where

$$\mathsf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{v}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \quad \mathsf{v}_3 = \begin{bmatrix} k \\ 1 \\ 1 \end{bmatrix}$$

(b) Set $k = 1$ and define the matrix $\mathsf{A} = [\mathsf{v}_1, \mathsf{v}_2, \mathsf{v}_3]$. Matrix A determines the linear transformation $\mathsf{A} : \mathbb{R}^3 \to \mathbb{R}^3$. Is this linear transformation injective (one-to-one)? Is it surjective (onto)?

SOLUTION:

(a) The volume of a parallelepiped formed with vectors v_1, v_2, and v_3 is the absolute value of the determinant of a matrix with these vectors as column vectors. In this case that determinant is given by

$$\begin{vmatrix} 1 & 3 & k \\ 2 & 2 & 1 \\ 3 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} - 3\begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} + k\begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} = 4 - 4k.$$

The volume of the parallelepiped is $V = 4$ iff holds

$$4 = |4 - 4k| \quad \Rightarrow \quad \begin{cases} 4 = 4 - 4k & \Rightarrow \quad \boxed{k = 0}, \\ 4 = -4 + 4k & \Rightarrow \quad \boxed{k = 2}. \end{cases}$$

(b) If $k = 1$, then $\det(\mathsf{A}) = 0$. This implies that the set $\{\mathsf{v}_1, \mathsf{v}_2\mathsf{v}_3\}$ is linearly dependent, so $N(\mathsf{A}) \neq \{\mathsf{0}\}$, which implies that A is **not injective**. This matrix is also **not surjective**, since the Nullity-Rank Theorem implies

$$3 = \dim N(\mathsf{A}) + \dim R(\mathsf{A}) \geqslant 1 + \dim R(\mathsf{A}) \quad \Rightarrow \quad \dim R(\mathsf{A}) \leqslant 2.$$

This establishes that $R(\mathsf{A}) \neq \mathbb{R}^3$. ◁

3. Determine whether the subset $V \subset \mathbb{R}^3$ is a subspace, where

$$V = \left\{ \begin{bmatrix} -a+b \\ a-2b \\ a-7b \end{bmatrix} \quad \text{with} \quad a, b \in \mathbb{R} \right\}.$$

If the set is a subspace, find an orthogonal basis in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

SOLUTION: Vectors in V can be written as follows,

$$\begin{bmatrix} -a+b \\ a-2b \\ a-7b \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} a + \begin{bmatrix} 1 \\ -2 \\ -7 \end{bmatrix} b$$

which implies that

$$V = \text{Span}\Big(\Big\{ \mathsf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \mathsf{v}_2 = \begin{bmatrix} 1 \\ -2 \\ -7 \end{bmatrix} \Big\} \Big) \quad \Rightarrow \quad \boxed{V \text{ is a subspace.}}$$

A basis for this subspace V is the set $\{\mathsf{v}_1, \mathsf{v}_2\}$ since these vectors are not collinear. This basis is not orthogonal though, because

$$\mathsf{v}_1 \cdot \mathsf{v}_2 = -10 \neq 0.$$

We construct and orthogonal basis for V by projecting v_2 onto v_1. That is, let us find the vector

$$\mathsf{v}_{2\perp} = \mathsf{v}_2 - \frac{\mathsf{v}_1 \cdot \mathsf{v}_2}{\|\mathsf{v}_1\|^2}\, \mathsf{v}_2 - \begin{bmatrix} 1 \\ -2 \\ -7 \end{bmatrix} - \frac{(-10)}{3} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\Big(\begin{bmatrix} 3 \\ -6 \\ -21 \end{bmatrix} + \begin{bmatrix} -10 \\ 10 \\ 10 \end{bmatrix} \Big) \quad \Rightarrow \quad \mathsf{v}_{2\perp} = \frac{1}{3} \begin{bmatrix} -7 \\ 4 \\ -11 \end{bmatrix}.$$

Since any non-zero vector proportional $\mathsf{v}_{2\perp}$ is orthogonal to v_1, an orthogonal basis for the subspace V is

$$\boxed{\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -7 \\ 4 \\ -11 \end{bmatrix} \right\}}.$$

◁

4. True or False: (Justify your answers.)

(a) If the set of columns of $\mathsf{A} \in \mathbb{F}^{m,n}$ is a linearly independent set, then $\mathsf{Ax} = \mathsf{b}$ has exactly one solution for every $\mathsf{b} \in \mathbb{F}^m$.

(b) The set of column vectors of an 5×7 is never linearly independent.

SOLUTION:

(a) **False.** The reason why this statement is false lies in the part of the sentence "for every $\mathsf{b} \in \mathbb{F}^m$". Do not get confused by the "exactly one solution" part. We now give an example of a system where the set of coefficient column vectors is linearly independent but the system has no solution. Take $\mathsf{A} \in \mathbb{R}^{3,2}$ and $\mathsf{b} \in \mathbb{R}^3$ given by,

$$\mathsf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

In this case $\mathsf{Ax} = \mathsf{b}$ has no solution.

(b) **True.** The biggest linearly independent set in $\mathbb{R}^5$ contains five vectors, so any set containing seven vectors is linearly dependent. ◁

5. Consider the vector space $\mathbb{R}^2$ with the standard basis $\mathcal{S}$ and let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation

$$[\boldsymbol{T}]_{ss} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}_{ss}.$$

Find $[\boldsymbol{T}]_{bb}$, the matrix of $\boldsymbol{T}$ in the basis $\mathcal{B}$, where $\mathcal{B} = \left\{ [\mathsf{b}_1]_s = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_s, [\mathsf{b}_2]_s = \begin{bmatrix} 1 \\ -2 \end{bmatrix}_s \right\}$.

SOLUTION: From the change of basis formulas we know that

$$[\boldsymbol{T}]_{bb} = \mathsf{P}^{-1}[\boldsymbol{T}]_{ss}\mathsf{P}, \qquad \mathsf{P} = \mathsf{I}_{bs} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}, \qquad \mathsf{P}^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}.$$

Therefore, we obtain

$$[\boldsymbol{T}]_{bb} = \frac{1}{4}\begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}_{ss}\begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} \quad \Rightarrow \quad \boxed{[\boldsymbol{T}]_{bb} = \frac{1}{2}\begin{bmatrix} 9 & -5 \\ 1 & -1 \end{bmatrix}}.$$

◁

6. Consider the linear transformations $\boldsymbol{T} : \mathbb{R}^3 \to \mathbb{R}^2$ and $\boldsymbol{S} : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\left[\boldsymbol{T}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{s_3}\right)\right]_{s_2} = \begin{bmatrix} x_1 - x_2 + x_3 \\ -x_1 + 2x_2 + x_3 \end{bmatrix}_{s_2}, \qquad \left[\boldsymbol{S}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{s_3}\right)\right]_{s_3} = \begin{bmatrix} 3x_3 \\ 2x_2 \\ x_1 \end{bmatrix}_{s_3},$$

where $\mathcal{S}_3$ and $\mathcal{S}_2$ are the standard basis of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively.

(a) Find a matrix $[\boldsymbol{T}]_{s_3 s_2}$ and the matrix $[\boldsymbol{S}]_{s_3 s_3}$. Show your work.

(b) Find the matrix of the composition $\boldsymbol{T} \circ \boldsymbol{S} : \mathbb{R}^3 \to \mathbb{R}^2$ in the standard basis, that is, find $[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2}$.

(c) Is $\boldsymbol{T} \circ \boldsymbol{S}$ injective (one-to-one)? Is $\boldsymbol{T} \circ \boldsymbol{S}$ surjective (onto)? Justify your answer.

SOLUTION:

(a) From the definition of the matrix of a linear transformation we get,

$$\boxed{[\boldsymbol{T}]_{s_3 s_2} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}}, \qquad \boxed{[\boldsymbol{S}]_{s_3 s_3} = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}}.$$

(b) Recalling the formula $[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2} = [\boldsymbol{T}]_{s_3 s_2}[\boldsymbol{S}]_{s_3 s_3}$, we obtain

$$[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 3 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \boxed{[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2} = \begin{bmatrix} 1 & -2 & 3 \\ 1 & 4 & -3 \end{bmatrix}}.$$

(c) We now obtain the reduced echelon form of matrix $[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2}$ to find out whether this composition is injective or surjective.

$$[\boldsymbol{T} \circ \boldsymbol{S}]_{s_3 s_2} = \begin{bmatrix} 1 & -2 & 3 \\ 1 & 4 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 3 \\ 0 & 6 & -6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

Since $\dim N(\boldsymbol{T} \circ \boldsymbol{S}) = 3 - \operatorname{rank}(\boldsymbol{T} \circ \boldsymbol{S}) = 1$, we conclude that $\boldsymbol{T} \circ \boldsymbol{S}$ is **not injective.** Since $\dim R(\boldsymbol{T} \circ \boldsymbol{S}) = 2 = \dim \mathbb{R}^2$ and $R(\boldsymbol{T} \circ \boldsymbol{S}) \subset \mathbb{R}^2$, we conclude that $R(\boldsymbol{T} \circ \boldsymbol{S}) = \mathbb{R}^2$, which shows that $\boldsymbol{T} \circ \boldsymbol{S}$ **is surjective.**. ◁

7. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}$ and the vector $\mathsf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$.

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the matrix equation $\mathsf{A}\mathsf{x} = \mathsf{b}$.

(b) Verify whether the vector $\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}$ belong to the space $R(\mathsf{A})^{\perp}$? Justify your answers.

SOLUTION:

(a) We find the normal equation $\mathsf{A}^T\mathsf{A}\hat{\mathsf{x}} = \mathsf{A}^T\mathsf{b}$ for $\hat{\mathsf{x}}$ and we solve it as follows.

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{A}^T\mathsf{A} = \begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix}.$$

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{A}^T\mathsf{b} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The normal equation is

$$\begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix} \hat{\mathsf{x}} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \hat{\mathsf{x}} = \frac{1}{(36-25)} \begin{bmatrix} 6 & -5 \\ -5 & 6 \end{bmatrix} 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \boxed{\hat{\mathsf{x}} = \frac{5}{11} \begin{bmatrix} 1 \\ 1 \end{bmatrix}}.$$

(b) We first compute the vector

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \frac{5}{11} \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \frac{5}{11} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} - \frac{11}{11} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} -12 \\ 4 \\ 4 \end{bmatrix}.$$

We conclude that

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) = \frac{4}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}.$$

Now it is simple to verify that $(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) \in R(\mathsf{A})^\perp$, since

$$\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = 0, \qquad \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 0 \quad \Rightarrow \quad \boxed{(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) \in R(\mathsf{A})^\perp}.$$

◁

8. Consider the matrix $\mathsf{A} = \begin{bmatrix} -1/2 & -3 \\ 1/2 & 2 \end{bmatrix}$.

(a) Show that matrix A is diagonalizable.

(b) Using that A is diagonalizable, find the $\lim_{k\to\infty} \mathsf{A}^k$.

SOLUTION:

(a) We need to compute the eigenvalues and eigenvectors of matrix A. The eigenvalues are the solutions of the characteristic equation

$$p(\lambda) = \begin{vmatrix} \left(-\frac{1}{2} - \lambda\right) & -3 \\ \frac{1}{2} & (2-\lambda) \end{vmatrix} = (\lambda - 2)\left(\lambda + \frac{1}{2}\right) + \frac{3}{2} = \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = 0.$$

We then obtain,

$$\boxed{\lambda_+ = 1}, \qquad \boxed{\lambda_- = \frac{1}{2}}.$$

The eigenvector corresponding to $\lambda_+ = 1$ is a non-zero solution of $\mathsf{A} - \mathsf{I}_2)\mathsf{v}^+ = \mathsf{0}$. The Gauss method implies,

$$\begin{bmatrix} -\frac{3}{2} & -3 \\ \frac{1}{2} & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^+ = -2v_2^+ \quad \Rightarrow \quad \boxed{\mathsf{v}^+ = \begin{bmatrix} -2 \\ 1 \end{bmatrix}}.$$

The eigenvector corresponding to $\lambda_- = 1/2$ is a non-zero solution of $\mathsf{A} - \mathsf{I}_2)\mathsf{v}^+ = \mathsf{0}$ and it is computed as follows,

$$\begin{bmatrix} -1 & -3 \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix} \to \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^- = -3v_2^- \quad \Rightarrow \quad \boxed{\mathsf{v}^- = \begin{bmatrix} -3 \\ 1 \end{bmatrix}}.$$

WE now know that matrix A is diagonalizable since $\{\mathsf{v}^+, \mathsf{v}^-\}$ is linearly independent. Introducing matrices P and D as follows,

$$\mathsf{P} = \begin{bmatrix} -2 & -3 \\ 1 & 1 \end{bmatrix}, \qquad \mathsf{D} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix},$$

we conclude that $\mathsf{A} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}$, that is,

$$\boxed{\mathsf{A} = \begin{bmatrix} -2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & -2 \end{bmatrix}}.$$

(b) Knowing that A is diagonalizable it is simple to compute $\lim_{k\to\infty} \mathsf{A}^k$. The calculation is

$$\begin{aligned} \lim_{k\to\infty} \mathsf{A}^k &= \lim_{k\to\infty} \begin{bmatrix} -2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \left(\frac{1}{2}\right)^k \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & -2 \end{bmatrix} \\ &= \begin{bmatrix} -2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & -2 \end{bmatrix} \quad \Rightarrow \quad \boxed{\lim_{k\to\infty} \mathsf{A}^k = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}}. \end{aligned}$$

◁

9. Let $\big(V, \langle\,,\rangle\big)$ be an inner product space with inner product norm $\|\ \|$. Let $\boldsymbol{T}: V \to V$ be a linear transformation and $\boldsymbol{x},\ \boldsymbol{y} \in V$ be vectors satisfying the following conditions:

$$\boldsymbol{T}(\boldsymbol{x}) = 2\,\boldsymbol{x}, \quad \boldsymbol{T}(\boldsymbol{y}) = -3\,\boldsymbol{y}, \quad \|\boldsymbol{x}\| = 1/3, \quad \|\boldsymbol{y}\| = 1, \quad \boldsymbol{x} \perp \boldsymbol{y}.$$

(a) Compute $\|\boldsymbol{v}\|$ for the vector $\boldsymbol{v} = 3\boldsymbol{x} - \boldsymbol{y}$.
(b) Compute $\|\boldsymbol{T}(\boldsymbol{v})\|$ for the vector $\boldsymbol{v}$ given above.

SOLUTION:
(a)

$$\begin{aligned} \|\boldsymbol{v}\|^2 &= \|3\boldsymbol{x} - \boldsymbol{y}\|^2 \\ &= \langle (3\boldsymbol{x} - \boldsymbol{y}), (3\boldsymbol{x} - \boldsymbol{y}) \rangle \\ &= 9\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 3\langle \boldsymbol{x}, \boldsymbol{y} \rangle - 3\langle \boldsymbol{y}, \boldsymbol{x} \rangle \\ &= 9\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 \\ &= 9\Big(\frac{1}{3}\Big)^2 + 1 \quad \Rightarrow \quad \boxed{\|\boldsymbol{v}\| = \sqrt{2}}. \end{aligned}$$

(b) We first compute $\boldsymbol{T}(\boldsymbol{v}) = \boldsymbol{T}(3\boldsymbol{x} - \boldsymbol{y}) = 3\,\boldsymbol{T}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{y})$. Recall that $\boldsymbol{T}(\boldsymbol{x}) = 2\boldsymbol{x}$, $\boldsymbol{T}(\boldsymbol{y}) = -3\boldsymbol{y}$, we conclude that $\boldsymbol{T}(\boldsymbol{v}) == 6\boldsymbol{x} - 3\boldsymbol{y}$. Then, it is simple to see that

$$\begin{aligned} \|\,\boldsymbol{T}(\boldsymbol{v})\|^2 &= \|6\boldsymbol{x} - 3\boldsymbol{y}\|^2 \\ &= \langle (6\boldsymbol{x} - 3\boldsymbol{y}), (6\boldsymbol{x} - 3\boldsymbol{y}) \rangle \\ &= 36\|\boldsymbol{x}\|^2 + 9\|\boldsymbol{y}\|^2 \\ &= 36\Big(\frac{1}{3}\Big)^2 + 9 \\ &= 4 + 9 \quad \Rightarrow \quad \boxed{\|\,\boldsymbol{T}(\boldsymbol{v})\| = \sqrt{13}}. \end{aligned}$$

◁

10. Consider the matrix $\mathsf{A} = \begin{bmatrix} 2 & -1 & 2 \\ 0 & 1 & h \\ 0 & 0 & 2 \end{bmatrix}$.

(a) Find all eigenvalues of matrix A and their corresponding algebraic multiplicities.

(b) Find the value(s) of the real number h such that the matrix A above has a two-dimensional eigenspace, and find a basis for this eigenspace.

SOLUTION:

(a) Since matrix A is upper triangular, its eigenvalues are the diagonal elements. Denoting by λ the eigenvalues and r their corresponding algebraic multiplicities we obtain,

$$\boxed{\lambda_1 = 2, \quad r_1 = 2}, \qquad \text{and} \qquad \boxed{\lambda_2 = 1, \quad r_2 = 1}.$$

(b) Since the eigenvalue $\lambda_2 = 1$ has algebraic multiplicity $r_2 = 1$, this eigenvalue cannot have a two-dimensional eigenspace, since $dim F_{\lambda_2} \leqslant r_2 = 1$. The only candidate is the eigenspace E_{λ_1}. Let us compute the non-zero elements in this eigenspace,

$$\mathsf{A} - 2\mathsf{I}_3 = \begin{bmatrix} 0 & -1 & 2 \\ 0 & -1 & h \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & -2 \\ 0 & 0 & h-2 \\ 0 & 0 & 0 \end{bmatrix}.$$

We look for the value of h such that the eigenspace E_{λ_1} is two-dimensional, which means that the reduced echelon form associated with the matrix above has rank equal to one. This means that this matrix has only one pivot, which in turns is a condition on h. The condition is

$$\boxed{h = 2}.$$

In this case, the eigenspace is given by the solutions of the homogeneous linear system with the coefficient matrix

$$\begin{bmatrix} 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_2 = -2x_3 \\ x_1, x_3 \quad \text{free.} \end{cases} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} x_3.$$

Therefore, a basis for E_{λ_1} is given by

$$\boxed{\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right\}}.$$

◁

Solutions to Practice Exam 2.

1. Consider the matrix $\mathsf{A} = \begin{bmatrix} -2 & 3 & -1 \\ 1 & 2 & -1 \\ -2 & -1 & 1 \end{bmatrix}$. Find the coefficients $(\mathsf{A}^{-1})_{13}$ and $(\mathsf{A}^{-1})_{21}$ of the inverse matrix of A. Show your work.

SOLUTION: The formula for the inverse matrix $\mathsf{A}^{-1} = \mathsf{C}^T / \det(\mathsf{A})$ implies that

$$\big(\mathsf{A}^{-1}\big)_{13} = \frac{C_{31}}{\det(\mathsf{A})} \qquad \big(\mathsf{A}^{-1}\big)_{21} = \frac{C_{12}}{\det(\mathsf{A})}.$$

We start computing the $\det(\mathsf{A})$, that is,

$$\begin{vmatrix} -2 & 3 & -1 \\ 1 & 2 & -1 \\ -2 & -1 & 1 \end{vmatrix} = -2 \begin{vmatrix} 2 & -1 \\ -1 & 1 \end{vmatrix} - 3 \begin{vmatrix} 1 & -1 \\ -2 & 1 \end{vmatrix} - \begin{vmatrix} 1 & 2 \\ -2 & -1 \end{vmatrix} = -2 + 3 - 3,$$

which gives us the result $\det(\mathsf{A}) = -2$. Now the cofactors,

$$C_{31} = (-1)^{(3+1)} \begin{vmatrix} 3 & -1 \\ 2 & -1 \end{vmatrix} = -1, \qquad C_{12} = (-1)^{(1+2)} \begin{vmatrix} 1 & -1 \\ -2 & 1 \end{vmatrix} = 1.$$

Therefore, we conclude that

$$\boxed{\big(\mathsf{A}^{-1}\big)_{13} = \frac{1}{2}}, \qquad \boxed{\big(\mathsf{A}^{-1}\big)_{12} = -\frac{1}{2}}.$$

◁

2. Consider the vector space $\mathbb{P}_3([0,1])$ with the inner product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\, dx.$$

Given the set $\mathcal{U} = \{\boldsymbol{p}_1 = x^2,\ \boldsymbol{p}_2 = x^3\}$, find an orthogonal basis for the subspace $U = \mathrm{Span}(\mathcal{U})$ using the Gram-Schmidt method on the set $\mathcal{U}$ starting with the vector $\boldsymbol{p}_1$.

SOLUTION: We call the orthogonal basis $\{\boldsymbol{q}_1,\ \boldsymbol{q}_2\}$. Stating with vector $\boldsymbol{p}_1$ means $\boldsymbol{q}_1 = \boldsymbol{p}_1$, hence

$$\boxed{\boldsymbol{q}_1(x) = x^2}.$$

The vector $\boldsymbol{q}_2$ is found with the formula

$$\boldsymbol{q}_2 = \boldsymbol{p}_2 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle}{\|\boldsymbol{q}_1\|^2}\, \boldsymbol{q}_1.$$

We need to compute the scalars

$$\|\boldsymbol{q}_1\|^2 = \int_0^1 x^4\, dx = \frac{x^5}{5}\Big|_0^1 \quad \Rightarrow \quad \|\boldsymbol{q}_1\|^2 = \frac{1}{5}.$$

$$\langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle = \int_0^1 x^5\, dx = \frac{x^6}{6}\Big|_0^1 \quad \Rightarrow \quad \langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle = \frac{1}{6}.$$

Therefore, we conclude that

$$\boxed{\boldsymbol{q}_2 = x^3 - \frac{5}{6}\, x^2}.$$

◁

3. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 3 & 1 & 1 \\ 2 & 6 & 3 & 0 \\ 3 & 9 & 5 & -1 \end{bmatrix}$.

(a) Verify that the vector $\mathsf{v} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ -2 \end{bmatrix}$ belongs to the null space of A.

(b) Extend the set $\{\mathsf{v}\}$ into a basis of the null space of A.

SOLUTION:

(a) The vector v belongs to $N(\mathsf{A})$, since

$$\mathsf{Av} = \begin{bmatrix} 1 & 3 & 1 & 1 \\ 2 & 6 & 3 & 0 \\ 3 & 9 & 5 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ -4 \\ -2 \end{bmatrix} = \begin{bmatrix} 3+3-4-2 \\ 6+6-12+0 \\ 9+9-20+2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \boxed{\mathsf{Av} = 0}.$$

(b) We first need to find a basis for the $N(\mathsf{A})$. So we find all solutions to the homogeneous equation $\mathsf{Ax} = 0$. We perform Gauss operations on matrix A;

$$\mathsf{A} \to \begin{bmatrix} 1 & 3 & 1 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 2 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 0 & 3 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} {}^3R \begin{cases} x_1 = -3x_2 - 3x_4, \\ x_3 = 2x_4. \end{cases}$$

Therefore, the solution x has the form

$$\mathsf{x} = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -3 \\ 0 \\ 2 \\ 1 \end{bmatrix} x_4 \quad \Rightarrow \quad \mathcal{U} = \left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\} \text{ is a basis of } N(\mathsf{A}).$$

The following calculations is useful to find a basis of $\mathbb{N}(\mathsf{A})$ containing v:

$$\begin{bmatrix} 3 & -3 & -3 \\ 1 & 1 & 0 \\ -4 & 0 & 2 \\ 2 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ -4 & 0 & 2 \\ 2 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & -1 \\ 0 & 2 & 1 \\ 0 & -4 & 2 \\ 0 & -2 & 3 \end{bmatrix}.$$

From the last matrix we know that the pivot columns are the first and second columns. That says that the set $\mathcal{N}$ below form a basis to the $N(\mathsf{A})$, where

$$\boxed{\mathcal{N} = \left\{ \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}}.$$

This set includes vector v. ◁

4. Use Cramer's rule to find the solution to the linear system

$$\begin{aligned} 2x_1 + x_2 - x_3 &= 0 \\ x_1 + x_3 &= 1 \\ x_1 + 2x_2 + 3x_3 &= 0. \end{aligned}$$

SOLUTION: If we write this system in the form $\mathsf{Ax} = \mathsf{b}$, then

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Following Cramer's rule, the components of the solution vector $\mathsf{x} = [x_i]$, for $i = 1, 2, 3$, are given by $x_i = \det(\mathsf{A}_i)/\det(\mathsf{A})$. Since

$$\det(\mathsf{A}) = \begin{vmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{vmatrix} = -(3-1) - 2(2+1) \quad \Rightarrow \quad \det(\mathsf{A}) = -8,$$

we obtain,

$$x_1 = \frac{1}{(-8)} \begin{vmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{vmatrix} = \frac{-(3+2)}{(-8)} \qquad \Rightarrow \qquad x_1 = \frac{5}{8};$$

$$x_2 = \frac{1}{(-8)} \begin{vmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{vmatrix} = \frac{(6+1)}{(-8)} \qquad \Rightarrow \qquad x_2 = -\frac{7}{8};$$

$$x_3 = \frac{1}{(-8)} \begin{vmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{vmatrix} = \frac{-(4-1)}{(-8)} \qquad \Rightarrow \qquad x_3 = \frac{3}{8}.$$

We conclude that

$$\boxed{\mathsf{v} = \frac{1}{8} \begin{bmatrix} 5 \\ -7 \\ 3 \end{bmatrix}}.$$

◁

5. Let $\mathcal{S}_3$ and $\mathcal{S}_2$ be standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively, and consider the linear transformation $\boldsymbol{T} : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$\left[\boldsymbol{T}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{s_3}\right)\right]_{s_2} = \begin{bmatrix} -x_1 + 2x_2 - x_3 \\ x_1 + x_3 \end{bmatrix}_{s_2},$$

and introduce the bases $\mathcal{U} \subset \mathbb{R}^3$ and $\mathcal{V} \subset \mathbb{R}^2$ given by

$$\mathcal{U} = \left\{ [\mathsf{u}_1]_{s_3} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_{s_3}, \; [\mathsf{u}_2]_{s_3} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_{s_3}, \; [\mathsf{u}_3]_{s_3} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}_{s_3} \right\},$$

$$\mathcal{V} = \left\{ [\mathsf{v}_1]_{s_2} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{s_2}, \; [\mathsf{v}_2]_{s_2} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}_{s_2} \right\}.$$

Find the matrices $[\boldsymbol{T}]_{s_3 s_2}$ and $[\boldsymbol{T}]_{uv}$. Show your work.

SOLUTION: Matrix $[\boldsymbol{T}]_{s_3 s_2} = \left[\boldsymbol{T}(\mathsf{e}_1)]_{s_2}, \; \boldsymbol{T}(\mathsf{e}_2)]_{s_2}, \; \boldsymbol{T}(\mathsf{e}_3)]_{s_2}\right]$, that is,

$$\boxed{[\boldsymbol{T}]_{s_3 s_2} = \begin{bmatrix} -1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}}.$$

The matrix $[\boldsymbol{T}]_{uv}$ can be computed with the change of basis formula

$$[\boldsymbol{T}]_{uv} = \mathsf{Q}^{-1} [\boldsymbol{T}]_{s_3 s_2} \mathsf{P}$$

where the change of basis matrices Q and P are given by

$$\mathsf{Q} = \mathsf{I}_{vs_2} = \begin{bmatrix} 1 & -3 \\ 2 & 2 \end{bmatrix} \quad \Rightarrow \quad \mathsf{Q}^{-1} = \mathsf{I}_{s_2 v} = \frac{1}{8} \begin{bmatrix} 2 & -2 \\ 3 & 1 \end{bmatrix}, \qquad \mathsf{P} = \mathsf{I}_{us_3} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Therefore, the matrix $[\boldsymbol{T}]_{uv}$ is given by

$$[\boldsymbol{T}]_{uv} = \frac{1}{8}\begin{bmatrix}2 & -2\\ 3 & 1\end{bmatrix}\begin{bmatrix}-1 & 2 & -1\\ 1 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}$$
$$= \frac{1}{8}\begin{bmatrix}2 & -2\\ 3 & 1\end{bmatrix}\begin{bmatrix}1 & -2 & 1\\ 1 & 2 & 1\end{bmatrix} \quad \Rightarrow \quad \boxed{[\boldsymbol{T}]_{uv} = \frac{1}{8}\begin{bmatrix}5 & 2 & 5\\ -1 & 6 & -1\end{bmatrix}}.$$

◁

6. Consider the inner product space $\left(\mathbb{R}^{2,2}, \langle\,,\,\rangle_F\right)$ and the subspace

$$W = \text{Span}\left\{\mathsf{E}_1 = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},\ \mathsf{E}_2 = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\right\} \subset \mathbb{R}^{2,2}.$$

Find a basis for $W^\perp$, the orthogonal complement of W.

SOLUTION: We first find the whole subspace $W^\perp$ and afterwards we find a basis. The matrix $\mathsf{X} \in W^{|\perp}$ iff $\langle \mathsf{E}_1, \mathsf{X}\rangle_F = 0$ and $\langle \mathsf{E}_2, \mathsf{X}\rangle_F = 0$. If we denote

$$\mathsf{X} = \begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix},$$

then the equations above have the form

$$0 = \langle \mathsf{E}_1, \mathsf{X}\rangle_F = \text{tr}\left(\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix}\right) = x_{21} + x_{12};$$
$$0 = \langle \mathsf{E}_2, \mathsf{X}\rangle_F = \text{tr}\left(\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix}\right) = x_{11} - x_{22}.$$

We can express the solutions in the following way,

$$x_{11} = x_{22}, \qquad x_{21} = -x_{12}.$$

Then, a general element $\mathsf{X} \in W^\perp$ has the form

$$\mathsf{X} = \begin{bmatrix}x_{22} & x_{12}\\ -x_{12} & x_{22}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} x_{22} + \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix} x_{12}$$

for arbitrary scalars x_{22}, x_{12}. So a basis of $W^\perp$ is given by

$$\boxed{\left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\right\}}.$$

◁

7. Consider the matrix $\mathsf{A} = \begin{bmatrix}1 & 2\\ 1 & -1\\ -2 & 1\end{bmatrix}$ and the vector $\mathsf{b} = \begin{bmatrix}0\\ 1\\ 0\end{bmatrix}$.

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the matrix equation $\mathsf{A}\,\mathsf{x} = \mathsf{b}$.

(b) Verify that the vector $\mathsf{A}\,\hat{\mathsf{x}} - \mathsf{b}$, where $\hat{\mathsf{x}}$ is the least-squares solution found in part (7a), belongs to the space $R(\mathsf{A})^\perp$, the orthogonal complement of $R(\mathsf{A})$.

SOLUTION:

(a) We first compute the normal equation $\mathsf{A}^T\mathsf{A}\hat{\mathsf{x}} = \mathsf{A}^T\mathsf{b}$. The coefficient matrix is

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix}1 & 1 & -2\\ 2 & -1 & 1\end{bmatrix}\begin{bmatrix}1 & 2\\ 1 & -1\\ -2 & 1\end{bmatrix} = \begin{bmatrix}6 & -1\\ -1 & 6\end{bmatrix};$$

while the source vector is

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & 1 & -2 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Therefore, the least-squares solution $\hat{\mathsf{x}}$ is obtained as follows,

$$\begin{bmatrix} 6 & -1 \\ -1 & 6 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{35} \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \boxed{\hat{\mathsf{x}} = \frac{1}{7} \begin{bmatrix} 1 \\ -1 \end{bmatrix}}.$$

(b) We first compute the vector $(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b})$, which is given by

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) = \begin{bmatrix} 1 & 2 \\ 1 & -1 \\ -2 & 1 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad (\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) = \frac{1}{7} \begin{bmatrix} -1 \\ -5 \\ -3 \end{bmatrix}.$$

It is simple to see that

$$\boxed{\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -5 \\ -3 \end{bmatrix} = -1 - 5 + 6 = 0}, \qquad \boxed{\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -5 \\ -3 \end{bmatrix} = -2 + 5 - 3 = 0}.$$

◁

8. Suppose that a matrix $\mathsf{A} \in \mathbb{R}^{3,3}$ has eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 4$.
(a) Find the trace of A, find the trace of A^2, and find the determinant of A.
(b) Is matrix A invertible? If your answer is "yes", then prove it and find $\det(\mathsf{A}^{-1})$; if your answer is "no", then prove it.

SOLUTION: Since Matrix A is 3×3 and has three different eigenvalues, then A is diagonalizable. Therefore, there exists an invertible matrix P such that $\mathsf{A} = \mathsf{P}\,\mathsf{D}\,\mathsf{p}^{-1}$, where $\mathsf{D} = \text{diag}[1, 2, 4]$.

(a)

$$\text{tr}\,(\mathsf{A}) = \text{tr}\left(\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}\right) = \text{tr}\left(\mathsf{P}^{-1}\mathsf{P}\,\mathsf{D}\right) = \text{tr}\left(\mathsf{D}\right) = 1 + 2 + 4 \quad \Rightarrow \quad \boxed{\text{tr}\,(\mathsf{A}) = 7}.$$

$$\text{tr}\,(\mathsf{A}^2) = \text{tr}\left(\mathsf{P}\,\mathsf{D}^2\,\mathsf{P}^{-1}\right) = \text{tr}\left(\mathsf{P}^{-1}\mathsf{P}\,\mathsf{D}^2\right) = \text{tr}\left(\mathsf{D}^2\right) = 1 + 4 + 16 \quad \Rightarrow \quad \boxed{\text{tr}\,(\mathsf{A}^2) = 21}.$$

$$\det(\mathsf{A}) = \det\left(\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}\right) = \det(\mathsf{P}^{-1})\,\det(\mathsf{D})\,\det(\mathsf{P}) = \frac{1}{\det(\mathsf{P})}\,\det(\mathsf{D})\,\det(\mathsf{P}) = \det(\mathsf{D}).$$

Since $\det(\mathsf{D}) = (1)(2)(4)$, we conclude that

$$\boxed{\det(\mathsf{A}) = 8}.$$

(b) Since matrix A has non-zero determinant, then it is invertible. We also know that $\det\left(\mathsf{A}^{-1}\right) = 1/\det(\mathsf{A})$, therefore

$$\boxed{\det\left(\mathsf{A}^{-1}\right) = \frac{1}{8}}.$$

◁

9. Consider the matrix $\mathsf{A} = \begin{bmatrix} 7 & 5 \\ 3 & -7 \end{bmatrix}$.
(a) Find the eigenvalues and eigenvectors of A.
(b) Compute the matrix e^{A}.

SOLUTION:
(a) The eigenvalues are the roots of the characteristic polynomial,

$$p(\lambda) = \begin{vmatrix} (7 - \lambda) & 5 \\ 3 & (-7 - \lambda) \end{vmatrix} = (\lambda - 7)(\lambda + 7) - 15 = \lambda^2 - 64.$$

Therefore, the eigenvalues are

$$\boxed{\lambda_+ = 8}, \qquad \boxed{\lambda_- = -8}.$$

The corresponding eigenvectors are the following: For $\lambda_+ = 8$ we obtain,

$$\mathsf{A} - 8\,\mathsf{I}_2 = \begin{bmatrix} -1 & 5 \\ 3 & -15 \end{bmatrix} \to \begin{bmatrix} 1 & -5 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^+ = 5v_2^+ \quad \Rightarrow \quad \boxed{\mathsf{v}^+ = \begin{bmatrix} 5 \\ 1 \end{bmatrix}}.$$

For $\lambda_- = -8$ we obtain

$$\mathsf{A} + 8\,\mathsf{I}_2 = \begin{bmatrix} 15 & 5 \\ 3 & 1 \end{bmatrix} \to \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad 3v_1^- = -v_2^- \quad \Rightarrow \quad \boxed{\mathsf{v}^- = \begin{bmatrix} 1 \\ -3 \end{bmatrix}}.$$

(b) Since $\mathsf{A} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}$, then $e^{\mathsf{A}} = \mathsf{P}e^{\mathsf{A}}\mathsf{P}^{-1}$, where

$$\mathsf{P} = \begin{bmatrix} 5 & 1 \\ 1 & -3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{P}^{-1} = \frac{1}{16}\begin{bmatrix} 3 & 1 \\ 1 & -5 \end{bmatrix}.$$

Therefore, it is simple to compute

$$\begin{aligned} e^{\mathsf{A}} &= \begin{bmatrix} 5 & 1 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} 8 & 0 \\ 0 & -8 \end{bmatrix}\frac{1}{16}\begin{bmatrix} 3 & 1 \\ 1 & -5 \end{bmatrix}, \\ &= \frac{1}{2}\begin{bmatrix} 5 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 1 & -5 \end{bmatrix}, \\ &= \frac{1}{2}\begin{bmatrix} 14 & 10 \\ 6 & -14 \end{bmatrix} \quad \Rightarrow \quad \boxed{e^{\mathsf{A}} = \begin{bmatrix} 7 & 5 \\ 3 & -7 \end{bmatrix}}. \end{aligned}$$

◁

10. Find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solution of the initial value problem

$$\frac{d}{dt}\mathsf{x}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where the matrix $\mathsf{A} = \begin{bmatrix} -5 & 2 \\ -12 & 5 \end{bmatrix}$ and the vector $\mathsf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

SOLUTION: We first need to compute the eigenvalues and eigenvectors of matrix A. The eigenvalues are the roots of the characteristic polynomial,

$$p(\lambda) = \begin{vmatrix} (-5-\lambda) & 2 \\ -12 & (5-\lambda) \end{vmatrix} = (\lambda+5)(\lambda-5) + 24 = \lambda^2 - 1.$$

Therefore, the eigenvalues are

$$\boxed{\lambda_+ = 1}, \qquad \boxed{\lambda_- = -1}.$$

The corresponding eigenvectors are the following: For $\lambda_+ = 1$ we obtain,

$$\mathsf{A} - \mathsf{I}_2 = \begin{bmatrix} -6 & 2 \\ -12 & 4 \end{bmatrix} \to \begin{bmatrix} -3 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad 3v_1^+ = v_2^+ \quad \Rightarrow \quad \boxed{\mathsf{v}^+ = \begin{bmatrix} 1 \\ 3 \end{bmatrix}}.$$

For $\lambda_- = -1$ we obtain

$$\mathsf{A} + \mathsf{I}_2 = \begin{bmatrix} -4 & 2 \\ 12 & 6 \end{bmatrix} \to \begin{bmatrix} -2 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad 2v_1^- = v_2^- \quad \Rightarrow \quad \boxed{\mathsf{v}^- = \begin{bmatrix} 1 \\ 2 \end{bmatrix}}.$$

Since the general solution to the differential equation above is

$$\mathsf{x}(t) = c_+\mathsf{v}^+ e^{\lambda_+ t} + c_-\mathsf{v}^- e^{\lambda_- t},$$

we conclude that

$$\mathsf{x}(t) = c_+ \begin{bmatrix} 1 \\ 3 \end{bmatrix} e^{t} + c_- \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-t}.$$

The initial condition implies that

$$\begin{bmatrix}1\\1\end{bmatrix} = \mathsf{x}_0 = \mathsf{x}(0) = c_+ \begin{bmatrix}1\\3\end{bmatrix} + c_- \begin{bmatrix}1\\2\end{bmatrix}.$$

We need to solve the linear system

$$\begin{bmatrix}1 & 1\\3 & 2\end{bmatrix}\begin{bmatrix}c_+\\c_-\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix}c_+\\c_-\end{bmatrix} = \frac{1}{(-1)}\begin{bmatrix}2 & -1\\-3 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix}c_+\\c_-\end{bmatrix} = \begin{bmatrix}-1\\2\end{bmatrix}.$$

So the solution to the initial value problem above is

$$\boxed{\mathsf{x}(t) = -\begin{bmatrix}1\\3\end{bmatrix} e^{t} + 2\begin{bmatrix}1\\2\end{bmatrix} e^{-t}}.$$

◁

Solutions to Practice Exam 3.

1. (a) Find the LU-factorization of the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & 2 \\ 2 & -8 & -3 \end{bmatrix}$.

(b) Use the LU-factorization above to find the solution of the linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$, where $\mathsf{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$.

Solution:

(a) We use Gauss operations to transform matrix A into upper triangular, as follows,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & 2 \\ 2 & -8 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & -12 & -9 \end{bmatrix} \to \boxed{\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 2 \\ 0 & 0 & 6 \end{bmatrix} = \mathsf{U}}.$$

The factors used to construct matrix U define the lower triangular matrix L as follows,

$$\boxed{\mathsf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix}}.$$

(b) We find the solution x of $\mathsf{A}\mathsf{x} = \mathsf{b}$ using the LU-factorization $\mathsf{A} = \mathsf{L}\mathsf{U}$ as follows:

$$\mathsf{L}\mathsf{U}\mathsf{x} = \mathsf{b} \quad \Leftrightarrow \quad \begin{cases} \mathsf{L}\mathsf{y} = \mathsf{b}, \\ \mathsf{U}\mathsf{x} = \mathsf{y}. \end{cases}$$

We first find vector y using forward substitution,

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{y} = \begin{bmatrix} 1 \\ -1 \\ -6 \end{bmatrix}.$$

We now solve for vector x using back substitution,

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -6 \end{bmatrix} \quad \Rightarrow \quad \boxed{\mathsf{x} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}}.$$

◁

2. Determine whether the following sets W_1 and W_2 are subspaces of the vector space $\mathbb{P}_2\big([0,1]\big)$. If your answer is "yes," find a basis of the subspace.

(a) $W_1 = \{\boldsymbol{p} \in \mathbb{P}_2\big([0,1]\big) \ : \ \int_0^1 \boldsymbol{p}(x)\,dx \leqslant 1 \ \}$;

(b) $W_2 = \{\boldsymbol{p} \in \mathbb{P}_2\big([0,1]\big) \ : \ \int_0^1 x\,\boldsymbol{p}(x)\,dx = 0 \ \}$.

Solution:

(a) W_1 is **not** a subspace. Take any element $\boldsymbol{p} \in W_1$ satisfying $\int_0^1 \boldsymbol{p}(x)\,dx < -1$. Multiply that element by (-1). The result is not in W_1, since $\int_0^1 -\boldsymbol{p}(x)\,dx > 1$.

(b) W_2 **is** a subspace. Proof: given two elements $\boldsymbol{p} \in W_2$, $\boldsymbol{q} \in W_2$, that is,

$$\int_0^1 x\,\boldsymbol{p}(x)\,dx = 0, \qquad \int_0^1 x\,\boldsymbol{q}(x)\,dx = 0,$$

then an arbitrary linear combination $(a\boldsymbol{p} + b\boldsymbol{q})$ also belongs to W_2, since

$$\int_0^1 x\,\big(a\boldsymbol{p}(x) + b\boldsymbol{q}(x)\big)\,dx = a\int_0^1 x\,\boldsymbol{p}(x)\,dx + b\int_0^1 x\,\boldsymbol{q}(x)\,dx = 0.$$

◁

3. Find a basis of $\mathbb{R}^4$ containing a basis of the null space of matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & -4 & 3 \\ 2 & 1 & 1 & 3 \\ 1 & 1 & -1 & 2 \\ 3 & 2 & 0 & 5 \end{bmatrix}.$$

SOLUTION: We first find a basis of $N(\mathsf{A})$ using the Gauss method,

$$\begin{bmatrix} 1 & 2 & -4 & 3 \\ 2 & 1 & 1 & 3 \\ 1 & 1 & -1 & 2 \\ 3 & 2 & 0 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & -4 & 3 \\ 0 & -3 & 9 & -3 \\ 0 & -1 & 3 & -1 \\ 0 & -4 & 12 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & -4 & 3 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore, a vector $\mathsf{x} = [x_i] \in \mathbb{R}^4$ belongs to $N(\mathsf{A})$ iff

$$\left.\begin{aligned} x_1 &= -2x_3 - x_4, \\ x_2 &= 3x_3 - x_4, \end{aligned}\right\} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} -2 \\ 3 \\ 1 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} x_4,$$

so a basis for $N(\mathsf{A})$ is the set

$$\mathcal{N} = \left\{ \begin{bmatrix} -2 \\ 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Now we can find a basis of $\mathbb{R}^4$ containing $\mathcal{N}$ as follows. We start with any basis of $\mathbb{R}^4$, say the standard basis, and we construct a matrix, B as follows:

$$\mathsf{B} = \begin{bmatrix} -2 & -1 & 1 & 0 & 0 & 0 \\ 3 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We then transform B into reduced echelon form using Gauss operations. The pivot positions indicate the vectors forming a linearly independent set:

$$\begin{bmatrix} -2 & -1 & 1 & 0 & 0 & 0 \\ 3 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ -2 & -1 & 1 & 0 & 0 & 0 \\ 3 & -1 & 0 & 1 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & -3 & 1 \end{bmatrix}.$$

Therefore, a basis of $\mathbb{R}^4$ containing $\mathcal{N}$ is the following,

$$\boxed{\left\{ \begin{bmatrix} -2 \\ 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}}.$$

◁

4. Given a matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, introduce its characteristic polynomial $p_A(\lambda) = \det(\mathsf{A} - \lambda\,\mathsf{I}_n)$. This polynomial has the form $p_A(\lambda) = a_0 + a_1\lambda + \cdots + a_n\lambda^n$ for appropriate scalars $a_0, \cdots, a_n$. Now introduce a matrix-valued function $P_A : \mathbb{F}^{n,n} \to \mathbb{F}^{n,n}$ as follows

$$P_A(\mathsf{X}) = a_0\,\mathsf{I}_n + a_1\mathsf{X} + \cdots + a_n\mathsf{X}^n.$$

Determine whether the following statement is true or false and justify your answer: If matrix $\mathsf{A} \in \mathbb{F}^{n,n}$ is diagonalizable, then $P_A(\mathsf{A}) = \mathsf{0}$. (That is, $P_A(\mathsf{A})$ is the zero matrix.)

SOLUTION: We start with a comment: $P_A(\mathsf{X}) \neq \det(\mathsf{A} - \mathsf{X}\,\mathsf{I})$. Notice that the left-hand side is a matrix while the right hand side is a scalar. So an argument saying that the statement is true because $\det(\mathsf{A} - \mathsf{A}) = 0$ is wrong. Once again, $P_A(\mathsf{A}) \in \mathbb{F}^{n,n}$ and $\det(\mathsf{A} - \mathsf{A}) = 0 \in \mathbb{F}$.

Having clarified that, let us start saying that matrix A is diagonalizable, that is, there exists an invertible matrix P and a diagonal matrix $\mathsf{D} = \operatorname{diag}[\lambda_1, \cdots, \lambda_n]$ such that $\mathsf{A} = \mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}$, where $\lambda_1, \cdots, \lambda_n$ are eigenvalues of matrix A. Therefore,

$$\begin{aligned} P_A(\mathsf{A}) &= P_A\big(\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1}\big) \\ &= a_0\,\mathsf{P}\,\mathsf{P}^{-1} + a_1\,\mathsf{P}\,\mathsf{D}\,\mathsf{P}^{-1} + \cdots + a_n\,\mathsf{P}\,\mathsf{D}^n\mathsf{P}^{-1} \\ &= \mathsf{P}\big(a_0\,\mathsf{I}_+ a_1\mathsf{D} + \cdots + a_n\,\mathsf{D}^n\big)\mathsf{P}^{-1} \\ &= \mathsf{P}\operatorname{diag}\big[p_A(\lambda_1), \cdots, p_A(\lambda_n)\big]\,\mathsf{P}^{-1}. \end{aligned}$$

Since $\lambda_1, \cdots, \lambda_n$ are eigenvalues of matrix A, we get that

$$p_A(\lambda_1) = 0, \;\cdots,\; p_A(\lambda_n) = 0.$$

We then conclude that $P_A(\mathsf{A}) = \mathsf{0}$. So the statement is **true.** ◁

5. Consider the vector space $\mathbb{R}^2$ with ordered bases

$$\mathcal{S} = \left(\mathsf{e}_{1s} = \begin{bmatrix}1\\0\end{bmatrix}_s, \mathsf{e}_{2s} = \begin{bmatrix}0\\1\end{bmatrix}_s\right), \qquad \mathcal{U} = \left(\mathsf{u}_{1s} = \begin{bmatrix}2\\1\end{bmatrix}_s, \mathsf{u}_{2s} = \begin{bmatrix}1\\2\end{bmatrix}_s\right).$$

Let $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation given by

$$[\boldsymbol{T}(\mathsf{u}_1)]_s = \begin{bmatrix}-3\\1\end{bmatrix}_s, \quad [\boldsymbol{T}(\mathsf{u}_2)]_s = \begin{bmatrix}1\\3\end{bmatrix}_s.$$

(a) Find the matrix T_{ss}.
(b) Find the matrix T_{uu}.

SOLUTION:
(a) From the data of the problem we get matrix

$$\mathsf{T}_{us} = \begin{bmatrix}-3 & 1\\ 1 & 3\end{bmatrix}.$$

We compute matrix T_{ss} with the change of basis formula

$$\mathsf{T}_{ss} = \mathsf{Q}^{-1}\mathsf{T}_{us}\mathsf{P}, \qquad \mathsf{P} = \mathsf{I}_{su}, \qquad \mathsf{Q} = \mathsf{I}_{ss}.$$

From the data of the problem we get

$$\mathsf{I}_{us} = \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \quad \Rightarrow \quad \mathsf{P} = \frac{1}{3}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}.$$

Therefore,

$$\mathsf{T}_{ss} = \begin{bmatrix}-3 & 1\\ 1 & 3\end{bmatrix}\frac{1}{3}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix} \quad \Rightarrow \quad \boxed{\mathsf{T}_{ss} = \frac{1}{3}\begin{bmatrix}-7 & 5\\ -1 & 5\end{bmatrix}}.$$

(b) We compute matrix T_{uu} with the change of basis formula

$$\mathsf{T}_{uu} = \mathsf{Q}^{-1}\mathsf{T}_{us}\mathsf{P}, \qquad \mathsf{P} = \mathsf{I}_{uu}, \qquad \mathsf{Q} = \mathsf{I}_{us} = \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}.$$

Therefore,

$$\mathsf{T}_{uu} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} -3 & 1 \\ 1 & 3 \end{bmatrix} \quad \Rightarrow \quad \boxed{\mathsf{T}_{uu} = \frac{1}{3}\begin{bmatrix} -7 & -1 \\ 5 & 5 \end{bmatrix}}.$$

◁

6. Consider the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix}$ and the vector $\mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the matrix equation $\mathsf{A}\mathsf{x} = \mathsf{b}$.
(b) Find the vector on $R(\mathsf{A})$ that is the closest to the vector b in $\mathbb{R}^3$.

SOLUTION:

(a) We first compute the normal equation $\mathsf{A}^T\mathsf{A}\hat{\mathsf{x}} = \mathsf{A}^T\mathsf{b}$. The coefficient matrix is

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix};$$

while the source vector is

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

Therefore, the least-squares solution $\hat{\mathsf{x}}$ is obtained as follows,

$$\begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 6 & -1 \\ -1 & 6 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad \Rightarrow \quad \boxed{\hat{\mathsf{x}} = \frac{2}{35}\begin{bmatrix} 11 \\ 4 \end{bmatrix}}.$$

(b) The vector in $R(\mathsf{A})$ that is the closest to b is the vector $\mathsf{b}_{\shortparallel} = \mathsf{A}\hat{\mathsf{x}}$, that is,

$$\mathsf{B}_{\shortparallel} = \mathsf{A}\hat{\mathsf{x}} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} = \frac{2}{35}\begin{bmatrix} 11 \\ 4 \end{bmatrix} \quad \Rightarrow \quad \boxed{\mathsf{b}_{\shortparallel} = \frac{2}{35}\begin{bmatrix} 19 \\ 18 \\ 15 \end{bmatrix}}.$$

We now check if the above result is true. If $\mathsf{b}_{\shortparallel}$ is correct, then $(\mathsf{b} - \mathsf{b}_{\shortparallel}) \in R(\mathsf{A})^{\perp}$. Let us verify this last condition: We first compute the vector $(\mathsf{b} - \mathsf{b}_{\shortparallel})$, which is given by

$$(\mathsf{b} - \mathsf{b}_{\shortparallel}) = \frac{1}{35}\left(\begin{bmatrix} 35 \\ 35 \\ 35 \end{bmatrix} - \begin{bmatrix} 38 \\ 36 \\ 30 \end{bmatrix}\right) \quad \Rightarrow \quad (\mathsf{b} - \mathsf{b}_{\shortparallel}) = \frac{1}{35}\begin{bmatrix} -3 \\ -1 \\ 5 \end{bmatrix}.$$

It is simple to see that

$$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ -1 \\ 5 \end{bmatrix} = -3 - 2 + 5 = 0, \qquad \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ -1 \\ 5 \end{bmatrix} = -6 + 1 + 5 = 0.$$

◁

7. Consider the inner product space $(\mathbb{C}^3, \cdot)$ and the subspace $W \subset \mathbb{C}^3$ given by

$$W = \text{Span}\left(\left\{\begin{bmatrix} i \\ 1 \\ i \end{bmatrix}\right\}\right).$$

(a) Find a basis for $W^{\perp}$, the orthogonal complement of W.
(b) Find an orthonormal basis for $W^{\perp}$.

SOLUTION:

(a) The vector $\mathsf{w} \in W^\perp$ iff holds

$$0 = \begin{bmatrix} i \\ 1 \\ i \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -i & 1 & -i \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = -iw_1 + w_2 - iw_3 \quad \Rightarrow \quad w_1 = -iw_2 - w_3.$$

Therefore,

$$\mathsf{w} = \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} w_2 + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} w_3.$$

A basis for $W^\perp$ is the set $\mathcal{U}$ given by

$$\boxed{\mathcal{U} = \left\{ \mathsf{u}_1 = \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}}.$$

(b) The basis above is not orthonormal, since

$$\begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} i & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = -i \neq 0.$$

We use the Gram-Schmidt method to find an orthonormal basis,

$$\mathsf{u}_{2\perp} = \mathsf{u}_2 - \frac{\mathsf{u}_1 \cdot \mathsf{u}_2}{\|\mathsf{u}_1\|^2} \mathsf{u}_1.$$

We already computed $\mathsf{u}_1 \cdot \mathsf{u}_2 = -i$, while we also need

$$\|\mathsf{u}_1\|^2 = \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} i & 1 & 0 \end{bmatrix} \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} = 2.$$

Then,

$$\mathsf{u}_{2\perp} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{(-i)}{2} \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \left(\begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} \right) \quad \Rightarrow \quad w_{2\perp} = \frac{1}{2} \begin{bmatrix} -1 \\ i \\ 2 \end{bmatrix}.$$

It is simple to see that an orthonormal basis of $W^\perp$ is given by

$$\boxed{\mathcal{V} = \left\{ \mathsf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}, \mathsf{v}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ i \\ 2 \end{bmatrix} \right\}}.$$

◁

8. (a) Find the eigenvalues and eigenvectors of matrix $\mathsf{A} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$, and show that A is diagonalizable.

(b) Knowing that matrix A above is diagonalizable, explicitly find a square root of matrix A, that is, find a matrix X such that $\mathsf{X}^2 = \mathsf{A}$. How many square roots does matrix A have?

SOLUTION:

(a) The eigenvalues are the roots of the characteristic polynomial,

$$p(\lambda) = \begin{vmatrix} (5-\lambda) & 4 \\ 4 & (5-\lambda) \end{vmatrix} = (\lambda - 5)^2 - 16 = \lambda^2 - 10\lambda + 9.$$

Therefore, the eigenvalues are computed by the well-known formula

$$\lambda_\pm = \frac{1}{2}\left(10 \pm \sqrt{100 - 36}\right) = \frac{1}{2}\left(10 \pm \sqrt{64}\right) = 5 \pm 4 \quad \Rightarrow \quad \boxed{\begin{array}{l} \lambda_+ = 9\,, \\ \lambda_- = 1\,. \end{array}}$$

The corresponding eigenvectors are the following: For $\lambda_+ = 9$ we obtain,

$$A - 9\,I_2 = \begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^+ = v_2^+ \quad \Rightarrow \quad \boxed{v^+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}.$$

For $\lambda_- = 1$ we obtain

$$A - I_2 = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^- = -v_2^- \quad \Rightarrow \quad \boxed{v^- = \begin{bmatrix} -1 \\ 1 \end{bmatrix}}.$$

Introducing the matrices

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}, \qquad P^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

it is simple to see that A is diagonalizable, that is,

$$\boxed{A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}.$$

(b) Since matrix X satisfies that $X^2 = A = P\,D\,P^{-1}$, matrix X^2 is diagonalizable, sharing the eigenvectors and eigenvalues with matrix A, that is,

$$X^2 = P \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} P^{-1},$$

We also know that a matrix X and its squared X^2 share their eigenvectors with the eigenvalues of the latter being the squares of the eigenvalues of the former, that is,

$$X = P \begin{bmatrix} X_{11} & 0 \\ 0 & X_{22} \end{bmatrix} P^{-1}, \quad \text{with} \quad x_{11} = \pm 3, \quad X_{22} = \pm 1.$$

The equation above says that there are four square-roots of matrix A. These matrices are the following,

$$\boxed{X_1 = P \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} P^{-1}}, \quad \boxed{X_2 = P \begin{bmatrix} -3 & 0 \\ 0 & 1 \end{bmatrix} P^{-1}},$$

$$\boxed{X_3 = P \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} P^{-1}}, \quad \boxed{X_4 = P \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} P^{-1}}.$$

◁

9. Consider the matrix $A = \begin{bmatrix} 8 & -18 \\ 3 & -7 \end{bmatrix}$.

(a) Find the eigenvalues and eigenvectors of A.

(b) Compute explicitly the matrix-valued function e^{At} for $t \in \mathbb{R}$.

SOLUTION:

(a) The eigenvalues are the roots of the characteristic polynomial,

$$p(\lambda) = \begin{vmatrix} (8-\lambda) & -18 \\ 3 & (-7-\lambda) \end{vmatrix} = (\lambda+7)(\lambda-8) + 54 = \lambda^2 - \lambda - 2.$$

Therefore, the eigenvalues are computed by the well-known formula

$$\lambda_\pm = \frac{1}{2}\big(1 \pm \sqrt{1+8}\big) = \frac{1}{2}(1 \pm 3) \quad \Rightarrow \quad \boxed{\begin{array}{l} \lambda_+ = 2\,, \\ \lambda_- = -1\,. \end{array}}$$

The corresponding eigenvectors are the following: For $\lambda_+ = 2$ we obtain,

$$A - 2\,I_2 = \begin{bmatrix} 6 & -18 \\ 3 & -9 \end{bmatrix} \to \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^+ = 3v_2^+ \quad \Rightarrow \quad \boxed{v^+ = \begin{bmatrix} 3 \\ 1 \end{bmatrix}}.$$

For $\lambda_- = -1$ we obtain

$$\mathsf{A} + \mathsf{I}_2 = \begin{bmatrix} 9 & -18 \\ 3 & -6 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^- = 2v_2^- \quad \Rightarrow \quad \boxed{\mathsf{v}^- = \begin{bmatrix} 2 \\ 1 \end{bmatrix}}.$$

(b) Since matrix A is diagonalizable, we know that

$$e^{\mathsf{A}t} = \mathsf{P}\, e^{\mathsf{D}t}\, \mathsf{P}^{-1}, \quad \text{where} \quad \mathsf{P} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathsf{D} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix},$$

that is,

$$\boxed{e^{\mathsf{A}t} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{2t} & 0 \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix}}.$$

◁

10. Find the function $\mathsf{x} : \mathbb{R} \to \mathbb{R}^2$ solution of the initial value problem

$$\frac{d}{dt}\mathsf{x}(t) = \mathsf{A}\,\mathsf{x}(t), \qquad \mathsf{x}(0) = \mathsf{x}_0,$$

where the matrix $\mathsf{A} = \begin{bmatrix} -7 & 12 \\ -4 & 7 \end{bmatrix}$ and the vector $\mathsf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

SOLUTION: We first need to compute the eigenvalues and eigenvectors of matrix A. The eigenvalues are the roots of the characteristic polynomial,

$$p(\lambda) = \begin{vmatrix} (-7-\lambda) & 12 \\ -4 & (7-\lambda) \end{vmatrix} = (\lambda+7)(\lambda-7) + 48 = \lambda^2 - 1.$$

Therefore, the eigenvalues are

$$\boxed{\lambda_+ = 1}, \qquad \boxed{\lambda_- = -1}.$$

The corresponding eigenvectors are the following: For $\lambda_+ = 1$ we obtain,

$$\mathsf{A} - \mathsf{I}_2 = \begin{bmatrix} -8 & 12 \\ -4 & 6 \end{bmatrix} \to \begin{bmatrix} 2 & -3 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad 2v_1^+ = 3v_2^+ \quad \Rightarrow \quad \boxed{\mathsf{v}^+ = \begin{bmatrix} 3 \\ 2 \end{bmatrix}}.$$

For $\lambda_- = -1$ we obtain

$$\mathsf{A} + \mathsf{I}_2 = \begin{bmatrix} -6 & 12 \\ -4 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad v_1^- = 2v_2^- \quad \Rightarrow \quad \boxed{\mathsf{v}^- = \begin{bmatrix} 2 \\ 1 \end{bmatrix}}.$$

Since the general solution to the differential equation above is

$$\mathsf{x}(t) = c_+ \mathsf{v}^+ e^{\lambda_+ t} + c_- \mathsf{v}^- e^{\lambda_- t},$$

we conclude that

$$\mathsf{x}(t) = c_+ \begin{bmatrix} 3 \\ 2 \end{bmatrix} e^t + c_- \begin{bmatrix} 2 \\ 1 \end{bmatrix} e^{-t}.$$

The initial condition implies that

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \mathsf{x}_0 = \mathsf{x}(0) = c_+ \begin{bmatrix} 3 \\ 2 \end{bmatrix} + c_- \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

We need to solve the linear system

$$\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} c_+ \\ c_- \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} c_+ \\ c_- \end{bmatrix} = \frac{1}{(-1)} \begin{bmatrix} 1 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} c_+ \\ c_- \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

So the solution to the initial value problem above is

$$\boxed{\mathsf{x}(t) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} e^t - \begin{bmatrix} 2 \\ 1 \end{bmatrix} e^{-t}}.$$

◁

References

[1] S. Hassani. *Mathematical physics*. Springer, New York, 2000. Corrected second printing.
[2] D. Lay. *Linear algebra and its applications*. Addison Wesley, New York, 2005. Third updated edition.
[3] C. Meyer. *Matrix analysis and applied linear algebra*. SIAM, Philadelphia, 2000.
[4] G. Strang. *linear algebra and its applications*. Brooks/Cole, India, 2005. Fourth edition.